Development of AFLP markers for *Haliotis midae* for linkage mapping

Daleen Badenhorst

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science at the University of Stellenbosch

Supervisor: Dr Rouvay Roodt-Wilding

March 2008

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any other university for a degree.

Signature:……

Date:……

ABSTRACT

Haliotis midae, is the only commercially important species of the six abalone species found in South African coastal waters and has become a lucrative commercial commodity. Wild stocks of *H. midae* are, however, no longer commercially sustainable due to a combination of environmental factors and poaching. The solution to the crisis is artificial production systems in the form of abalone farms. An abalone enhancement programme was initiated in South Africa in 2006, funded by industry and government. This programme focuses on the elucidation of the abalone genome and genetic factors contributing to increased productivity, thereby aiding the commercial production of abalone.

The aims of this study, the first of its kind concerning *H. midae*, were to develop AFLP-based markers (specifically fluorescent AFLP analysis); to monitor the segregation of these markers in a single full-sib family and to use the markers and additional microsatellite markers to generate the first preliminary linkage map for *H. midae*.

Genomic DNA of sufficient quality and purity for fluorescent AFLP analysis was obtained from 3.5-month-old *H. midae* juveniles. Preliminary linkage maps were constructed using AFLP and microsatellite markers segregating in an F_1 family following a pseudo-testcross mapping strategy. Twelve AFLP primer combinations, producing 573 segregating peaks, and 10 microsatellite markers were genotyped in the parents and 108 progeny of the mapping family. Of the 573 segregating AFLP peaks genotyped, 241 segregated in a 1:1 ratio and 332 in a 3:1 ratio. Of these AFLP markers, 90 segregated according to the expected 1:1 Mendelian ratio and 164 segregated according to the expected 3:1 Mendelian ratio at the $P = 0.05$ level and were used for linkage analysis. Of the 10 microsatellite markers genotyped, nine were informative for linkage mapping analysis.

Preliminary male and female genetic linkage maps were developed using markers segregating in the female or male parent. A total of 12 and 10 linkage groups were detected for the female and male maps respectively. The female map covered 1473.5cM and consisted of 56 markers, and the male map covered 738.9cM consisting of 30 markers. Markers with segregation distortion were observed as previously reported in other abalone species and potential homology between one of the linkage groups of the

male map and two of the linkage groups of the female map were identified using the 3:1 segregating AFLP markers.

In conclusion, the genetic linkage map presented here, despite the fact that it has relatively low genome coverage and low marker density, forms an ideal starting point for more detailed study of the *H. midae* genome and will provide a scaffold for basic and applied studies in abalone. A high-density linkage map of *H. midae* should in future be developed with additional co-dominant molecular markers, such as microsatellites, to improve the transferability of the linkage map between different laboratories and among populations. A high-density linkage map will facilitate the mapping of QTL of commercially important traits (i.e. growth) and future MAS breeding programmes.

OPSOMMING

Perlemoenspesie, *Haliotis midae*, is die enigste spesie van kommersiële belang van die ses wat in die kuswater van Suid-Afrika aangetref word en het ‘n winsgewende handelskommoditeit in Suid-Afrika geword. Die ontginning van natuurlike *H. midae* populasies is egter, as gevolg van ‘n kombinasie van omgewingsfaktore en stropery nie meer kommersieel volhoubaar nie. Die perlemoenkrisis kan die hoof gebied word deur kunsmatige produksiesisteme op perlemoenplase tot stand te bring. ‘n Perlemoen verbeteringsprogram is in 2006 in Suid-Afrika geïnisieer en word deur die industrie en regering befonds. Die program focus op die ontrafeling van die perlemoen genoom en die genetiese faktore wat bydrae tot verhoogde produksie. Sodanige inligting kan gebruik word om kommersiële perlemoenproduksie te bevorder.

Die doel van hierdie studie, die eerste met *H. midae*, is om AFLP-gebaseerde merkers (spesifiek fluoresserende AFLP analise) te ontwikkel; die segregasie van hierdie merkers te monitor in ‘n enkel volledige verwante familie en die merkers en addisionele mikrosatelliet merkers te gebruik om die eerste voorlopige koppelingskaart vir *H. midae* te genereer.

Genomiese DNS van genoegsame kwaliteit en suiwerheid vir fluoresserende AFLP analise is ge-ekstraeer uit 3.5-maand-oue *H. midae* individue. Voorlopige koppelingskaart is gekonstrueer deur van segregerende AFLP en mikrosatelliet merkers in ‘n F_1 familie gebruik te maak deur ‘n pseudo-kruistoets karteringstrategie te volg. Twaalf AFLP inleier kombinasies, wat 573 segregerende fragmente geproduseer het, en 10 mikrosatelliet merkers is gegenotipeer in die ouers en 108 individue van die nageslag van die karteringsfamilie. Van die 573 segregerende AFLP merkers wat gegenotipeer is, het 241 in ‘n 1:1 verhouding en 332 in ‘n 3:1 verhouding gesegregeer. Van hierdie AFLP merkers, het 90 volgens die verwagte 1:1 Mendeliese verhouding en 164 volgens die 3:1 Mendeliese verhouding by die $P = 0.05$ gesegregeer vlak en is vir die koppelingsanalise gebruik. Van die 10 mikrosatelliet merkers gegenotipeer, was 9 informatief vir koppeling karteringsanalise.

Voorlopige manlike en vroulike genetiese koppelingskaarte is ontwikkel met gebruik te maak van merkers wat in die manlike of vroulike ouer segregeer het. ‘n Totaal van 12 en 10 koppelingsgroepe is onderskeidelik in die vroulike en manlike karate gegenereer. Die

vroulike kaart dek 1473.5cM and bestaan uit 56 merkers, terwyl die manlike kaart 738.9cM beslaan het met 30 merkers. Merkers wat segregasie distorsie toon is waargeneem soos voorheen in ander perlemoenspesies gerapporteer. Potensiële ooreenstemming tussen een van die koppelingsgroepe van die manlike kaart en twee van die koppelingsgroepe van die vroulike kaart is aangetoon deur van die 3:1 segregerende AFLP merkers gebruik te maak.

Die genetiese koppelingskaarte verskaf wel ‘n relatiewe lae genoomdekking en ‘n lae merkerdigtheid, maar is ‘n ideale vertrekpunt vir meer gedetailleerde studie van die *H. midae* genoom en dien as ‘n raamwerk vir toekomstige basiese en toegepaste studies in perlemoennavorsing. ‘n Hoëdigtheid koppelingskaart van *H. midae* moet in die toekoms ontwikkel word met gebruik van bykomstige ko-dominante molekulêre merkers, soos mikrosatelliete. Dit sal die oordraagbaarheid van die koppelingskaart tussen verskillende laboratoria asook tussen populasies verbeter. ‘n Hoëdigtheid koppelingskaart sal die kartering van kwantitatiewe kenmerk loki (KKL) vir kommersieel belangrike kenmerke (onder andere groeikrag) en toekomstige merker bemiddelde seleksie (MBS) teelprogramme moontlik maak.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. Rouvay Roodt-Wilding for her faithful encouragement, assistance, input and comments during the study and the preparation of this manuscript. Also special thanks to Carel van Heerden for his enthusiasm and technical assistance, especially with the Genemapper 3.0 software and Aletta van der Merwe for her advice and help throughout the project, as well as Ruhan Slabbert for fruitful discussions regarding DNA extraction methodologies. Lize Schoonbee of Irvin and Johnson abalone hatchery (Danger Point, South Africa) is thanked for supplying abalone larvae and small juveniles. Thanks go to Stellenbosch University for providing research facilities and finally the South African abalone industry, as well as the Innovation Fund, is thanked for providing funding for this research and monetary support in the form of a bursary.

TABLE OF CONTENTS

ABBREVIATIONS

α	alpha
ABI	Applied Biosystems
AFLP	amplified fragment length polymorphism
ATP	adenosine triphosphate
β	beta
BAC	bacterial artificial chromosome
bp	base pairs
BSA	bovine serum albumin
χ^2	chi-square
cm	centimeter
cM	centiMorgan
C_{oa}	total map coverage
C_{of}	framework map coverage
CTAB	N-cetyl-N, N, N-trimethyl ammonium bromide
^{o}C	degrees Celsius
df	degrees of freedom
ddH_20	double distilled water
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
*Eco*RI	restriction endonuclease obtained from the bacteria *Escherichia coli*
EDTA	ethylene diamine tetra-acetate
EST	expressed sequence tag
et al.	and company

F_1	first generation (progeny)
F_2	second generation (progeny)
FAM	blue ABI-fluorescent label; 5 - carboxylfluorescein
G	estimated genetic length
G_e	average estimated genome length
G_{e1}	estimated genome length (Fishman *et al.,* 2001)
G_{e2}	estimated genome length (Chakravarti *et al*., 1991)
G_{oa}	total length of map
G_{of}	length of framework map
H_o	observed heterozygosity
H_e	expected heterozygosity
Ho	null hypothesis
HCl	hydrochloric acid
JOE	green ABI-fluorescent label
kb	kilobase
KCl	potassium chloride
kg	kilogram
LG	linkage group
LOD	log of the odds (statistical measure of significance)

M	molar
m/v	mass to volume
MAS	marker-assisted selection
mg	milligram
$MgCl_2$	magnesium chloride
min	minute
ml	millilitre
mm	millimeter
mM	millimolar
MS	microsatellite
*Mse*I	restriction endonuclease isolated from an *E. coli* strain that carries the *Mse*I gene from the *Micrococcu*s species
NaCl	sodium chloride
NED	yellow ABI-fluorescent label
ng	nanogram
%	percentage
P	probability
PAA	poly-acrylamide gel
PCR	polymerase chain reaction
PET	red ABI-fluorescent label
pg	picogram
pH	concentration of hydrogen ions in a solution expressed traditionally as its pH
pmol	picomole
QTL	quantitative trait loci

r	*P*earson's *r*
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
S	second
SDL	segregation distortion loci
SNP	single nucleotide polymorphism
SSR	simple sequence repeat
TAC	total allowable catch
Taq	*Thermus aquaticus* DNA polymerase
TE	tris, EDTA
Tris	2-amino-2-(hydroxymethyl)-1, 3-proanediol
TURF	territorial user rights in fisheries
U	unit
µg	microgram
µl	microlitre
µM	micromolar
UV	ultra-violet
v/v	volume to volume
VIC	green ABI-fluorescent label
YAC	yeast artificial chromosome

LIST OF FIGURES

CHAPTER 1 – INTRODUCTION

CHAPTER 2 - DNA EXTRACTIONS

CHAPTER 3 - AFLP MARKER DEVELOPMENT AND ANALYSIS

CHAPTER 4 - MICROSATELLITE MARKER ANALYSIS

CHAPTER 5 - LINKAGE MAPPING

LIST OF TABLES

CHAPTER 2 – DNA EXTRACTIONS

CHAPTER 3 - AFLP MARKER DEVELOPMENT AND ANALYSIS

CHAPTER 4 - MICROSATELLITE MARKER ANALYSIS

CHAPTER 5 - LINKAGE MAPPING

LIST OF TABLES

CHAPTER 1

INTRODUCTION

CHAPTER 1 - INTRODUCTION

1.1. Background

Abalone, *Haliotis midae*, has become a lucrative commercial commodity in South Africa with artificial production systems (abalone farms) being the main contributors to the overseas export market. Efforts to improve productivity include determining the effects of environmental and genetic factors separately, and in combination, in order to determine optimal production criteria.

Studies of a genome of any species are important for the genetic improvement of that species. Genetic improvement entails the manipulation of the natural genetic variation found within a specific species to increase cultured production. The aim of such programmes is therefore the exploitation of the intrinsic biological potential of a cultured species. To date very few genetic studies have been done on the South African cultured abalone, *H. midae*. Various genetic and molecular techniques, for example microsatellites and amplified fragment length polymorphisms (AFLPs), can be applied in investigating the natural genetic variation present in abalone, with the eventual aim of improvement of *H. midae* and commercial abalone farms through genetic improvement programmes.

Decisions regarding which main traits are of commercial and economic importance (for example growth) are made by the local abalone industry and these traits are then targeted. Once the genome has been elucidated genetic programmes can focus on these characteristics and thereby ensure production gain. Increased worldwide demand for abalone products has increased market competition. Therefore, for the South African industry to maintain their competitive edge, they have to examine improvement programmes and specifically genetic improvement programmes as the logical next step and low-risk solution.

An abalone enhancement programme was initiated in South Africa in 2006 funded by industry and government, which focuses on genetic factors contributing to increased productivity. Genetic information will assist in the identification of abalone with higher survival rates and/or fast growers at an early stage when phenotypic observations are imprecise, thereby benefiting the commercial abalone farmers. For example, identification of faster growing abalone or animals with improved disease resistance at an early stage will

reduce production time and cost for market size animals (currently exported only at 3-4 years of age).

To prevent the extinction of dwindling wild abalone populations in South African coastal waters, all commercial fishing rights will be suspended indefinitely from February 2008. The abalone enhancement programme is consequently invaluable for maintaining genetic diversity of farmed abalone, as broodstock can no longer be collected from the wild. Increased abalone farm productivity is imperative, because when the South African industry can meet the worldwide demand for abalone products; it will lead to a reduction in the demand for poached abalone. This will alleviate the pressure on wild abalone populations and allow wild abalone populations to recover.

1.2. Abalone

1.2.1. Classification

Mollusca, Gastropoda, Prosobranchia (Winkles, limpets and whelks), Haliotidae (ear shaped shells), *Haliotis*, *Haliotis midae* (Fishtech, Inc, online).

1.2.2. Anatomy

Abalone are gastropod molluscs (Fallu, 1991), considered to be large herbivorous marine snails (Britz, 1991). Sketches of the abalone anatomy are shown in Figure 1.1 and images of mature abalone in Figure 1.2. The shell of an abalone covers the exposed area of the animal, with only the posterior of the animal remaining unprotected (Fallu, 1991). To a non-biologist the shell appears to be half a bivalve shell; namely a univalve. Careful assessment shows that there is a great deal of torsion of the shell, despite the fact that the shell is relatively flat (Landau, 1992) with an oval shape. The shell is similar to the common snail as it is in the form of a spiral, but due to its flattened shape the spiral is more visible from underneath. The head of the abalone is located at the anterior end (Fallu, 1991).

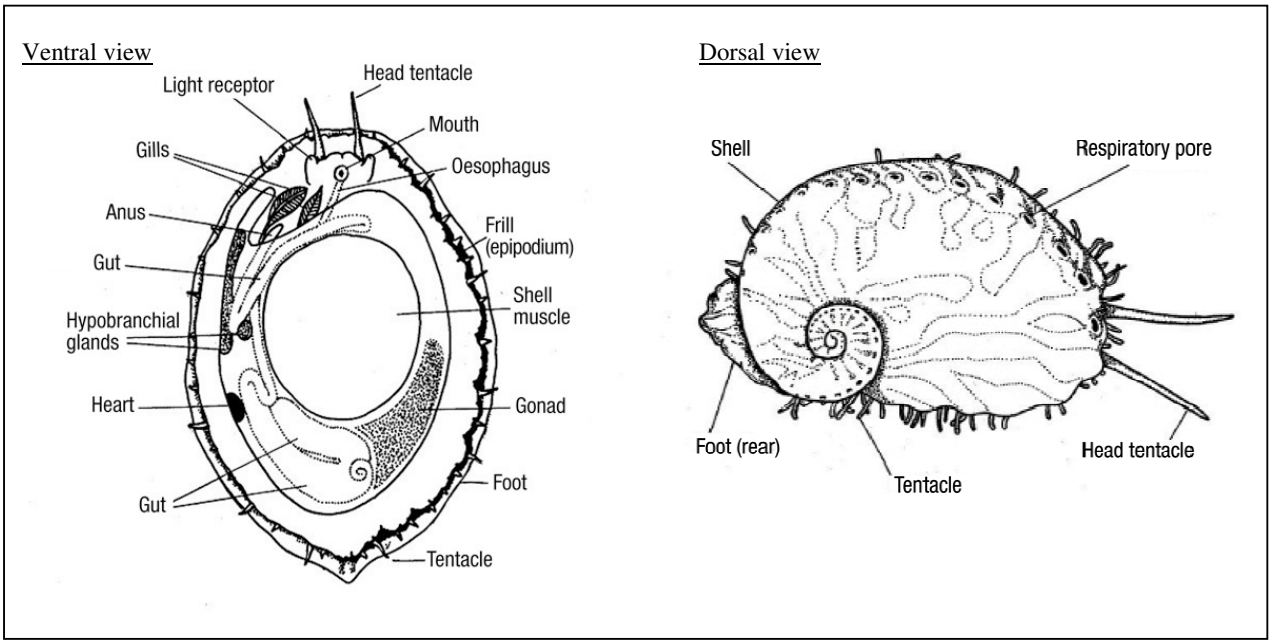
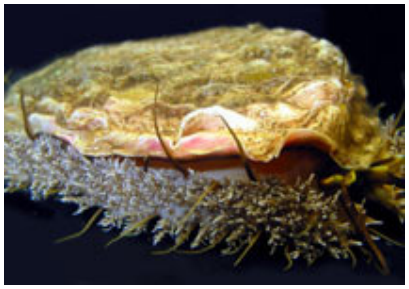


Figure 1.1. Ventral and dorsal view of the anatomy of the abalone (Fallu, 1994 [online]).

Ferrario, M (online)



Tony and Lisette Lewis Foundation (Online)

Sweijd, N (online)

Figure 1.2. Images of mature abalone.

The shell contains a row of holes, with the anterior holes being the largest, while the smaller holes at the posterior are frequently blocked due to their diminutive size (Fallu, 1991). These holes serve as areas for respiration, removal of waste and the liberation of gametes (Landau, 1992). The gills, for respiration, are located underneath these holes in the mantle cavity. It is in the mantle cavity that the other functions besides respiration are carried out, as the cavity houses the gills, excretory and digestive systems and the reproductive system which also release into it (Fallu, 1991).

Growth of the shell occurs by the accumulation of new material at the anterior side. The shell grows in a spiral, by rotating on the body, as the head of the animal must point forward at all times. The length of the shell is generally a good indicator of the size and weight of the abalone (Fallu, 1991).

The foot of the abalone is used to hold onto surfaces, such as a seabed and is similar to a snail in terms of how they grip surfaces. The part of the foot that can be seen when the animal is inverted is called lips. These lips have a tough outer layer that helps protect against predators. Tentacles are found on the lip, their function most likely being to sense and taste. The foot comprises the sought after and commercially prized and lucrative foot muscle (Fallu, 1991).

The head, containing the mouth, is located at the front of the foot. The tentacles found here are much larger than the tentacles found on the lips and are similar in shape to the land snails' eyestalks (Fallu, 1991).

1.2.3. Diet, habitat and movement

Abalone are herbivores and nocturnal feeders, a strategy thought to be essential in predator-avoidance (Wood and Buxton, 1996a). Abalone are similar to many gastropod molluscs in that they display a slow growth rate due to their fairly ineffective ability to convert their herbivorous diet into animal tissue (Nash, 1991). They appear to dislike light and during the daytime will be found hiding under rocky overhangs or in dark crevices (Fallu, 1991) where they feed primarily by trapping drift seaweed beneath their muscular foot. This species may intermittently graze on plants in close proximity or forage (Wood and Buxton, 1996a) and may feed on a wide range of algae as well, although they appear to prefer kelp (Barkai and Griffiths, 1986).

Abalone, furthermore, tend to remain in the same position if they have found a satisfactory feeding site. Some abalone have been known to remain in the exact same spot for several years (Wood and Buxton, 1996a). A stimulus to feed is generally initiated by the energetic movement of the surrounding water, as these conditions protect them more from predators and increases the likelihood of seaweed being washed past in close proximity (Fallu, 1991). Their preferred habitats are thus coastal rocky shores found at the low water mark (Barkai and Griffiths, 1986).

Abalone larvae are planktonic and do not feed as they lack the necessary equipment (Fallu, 1991). Newly settled juveniles will feed on diatoms (mixture of micro-algae and bacteria) scraped off the rock surfaces (Wood and Buxton, 1996a).

1.2.4. Distribution

There are six species of abalone in South Africa, but of the six only one, *H. midae* (Britz, 1991), better known as "perlemoen," is of commercial importance. The genus *Haliotis* consists of approximately fifty-six different species located worldwide (Geiger, 2000). *Haliotis midae* is distributed from St. Helena Bay (west coast) to Port St. Johns (east coast) (Fielding, 1995). The distribution of the six abalone species in South Africa is shown in Figure 1.3.

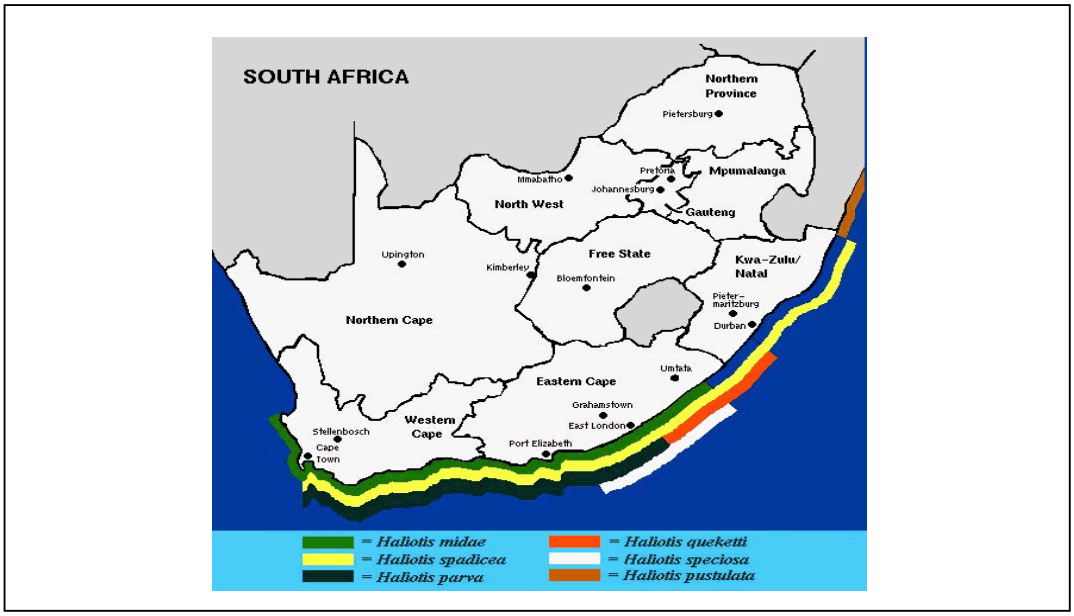


Figure 1.3. The distribution of South African abalone species (South African Abalone, online).

1.2.5. Life cycle

The life cycle (Figure 1.4) begins with spawning when sperm and eggs are liberated from the mature abalone. External fertilisation results in a zygote, with several divisions leading to the development of a larva from the fertilised egg. The larvae develop and change through a trochophore stage (± 12-24 hours) followed by veliger stage (± 1 week), after which they sink and settle on the seabed (termed settlement). The abalone body then starts to convert to a miniature copy of the adult and at this stage is known as a juvenile. The abalone continues to grow even after becoming sexually mature (Fallu, 1991) and *H. midae* can reach a shell length of up to 230mm (Steinberg, 2005 [online]).

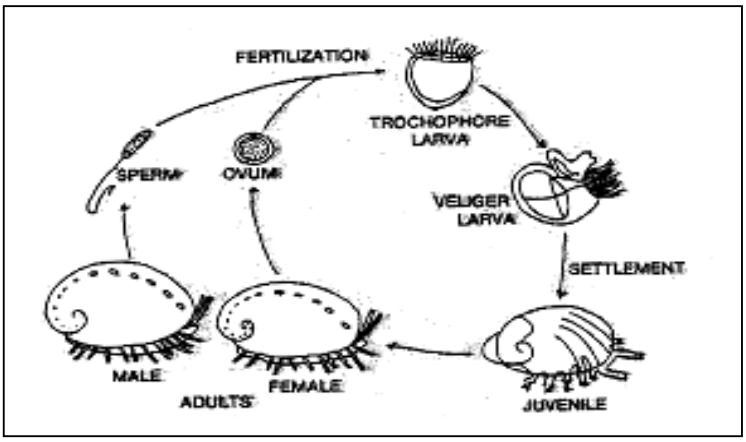


Figure 1.4. Life cycle of abalone (Fishtech, Inc, online).

1.2.6. Reproduction

The sex of the animals can be determined by examining the colour of their gonads, located by moving the foot away from the shell to the right hand side of the body. The ovaries are brown, dark green, gray or violet, while the testes are white, pink, yellow or a light brown or green (Landau, 1992). The eggs are usually a greenish colour and the sperm has a white appearance (Fallu, 1991).

Haliotis midae have been shown to have an extended breeding season, the main period being between April and July, during autumn and winter (Wood and Buxton, 1996b).

1.2.7. Resource under threat

Haliotis midae is the only South African abalone species that is of commercial importance (Sweijd *et al.*, 1998). Their sessile nature and easy accessibility in shallow coastal areas, including their high market value make them an ideal and prime target for part-time fishermen and poachers (Nash, 1991). Wild abalone stocks were well managed and plentiful until the early 1990s. Since then, the combination of environmental factors and poaching has devastated wild stocks to the point that commercial stocks are presently unsustainable (Hauck and Sweijd, 1999).

The gravity of the situation was realised in the late 1990s and this resulted in a total allowable catch (TAC) being set at 693 tons in 2000. The TAC for the 2002/2003 season was lowered further to 430 tons, while the illegal harvesting of abalone in 2001 was approximately 1023 tons. From 2003, only legal entities, such as abalone processing factories and certain divers could obtain commercial abalone rights and the rights of large fishing factories were to be phased out over a three year period (Steinberg, 2005 [online]). Further action was taken in 2004 when a new management approach was instated; Territorial User Rights in Fisheries (TURF). The TURF approach incorporates elements of co-management, which aim to address illegal harvesting and aid in the restoration of the overexploited abalone resource (Plagányi and Butterworth, 2004). In 2006/7 the legal commercial quotas reached an all time low of 125 tons (Macleod, 2007 [online]) and in October 2007, South African marine authorities took drastic measures to protect the dwindling wild abalone populations from commercial extinction and instated an indefinite postponement of all abalone fishing in South African waters, effective from February 2008 (Benton, 2007 [online]).

1.2.7.1. Poaching

Poaching has affected the legitimate industries to such an extent that some abalone fisheries around the world have collapsed. South African fisheries are especially under threat, as the reduced profit caused by poaching adversely impacts the development of abalone farms. At present the commercial fishery is unsustainable, as predicted by Steinberg (2005 [online]). Estimates have shown that at least 40 tons of *H. midae*, worth roughly US $1 million, are poached annually and the above figure is predicted to increase each year (Sweijd *et al*., 1998). In South Africa, the quantity of illegal abalone confiscated has increased dramatically from 21000 in 1994 to more than 1 million in 2006 (Roelf, 2007 [online]). Furthermore, 60% of the abalone confiscated from poachers in 2007 was undersize (Macleod, 2007 [online]).

Abalone poaching has increased in recent years in South Africa with the establishment of poaching syndicates, which illegally obtain, process and export the abalone to the Far East. In the eastern market *H. midae* is in great demand (Sweijd *et al*., 1998), fetching between $22 and $38 per kg (Roelf, 2007 [online]), thereby making abalone poaching a lucrative business. In addition, shells have an ornamental value and are highly prized (Nash, 1991).

1.2.7.2. Environmental factors: Rock Lobster

Adult abalone have virtually no predators besides man, while juvenile abalone and larvae are vulnerable to predation by numerous animals, such as octopuses, starfish, whelks and crabs. A result of this predation is that juvenile abalone hide in and under dark crevices until they possess the same shell strength as the adult abalone, at which point they can move out of hiding (Tarr *et al*., 1996). An unexplained invasion to the south coast of West Coast rock lobster, *Jasus ialandii*, has had a markedly negative impact on abalone populations in the area (Mayfield and Branch, 2000). The diet of the rock lobster consists of sea urchins, *Parechinus angulosus,* under which juvenile abalone shelter while their shells are not yet hard enough to protect them from predators. Studies have found a positive relationship between juveniles of *H. midae* and sea urchins, as both variables (*H. midae* and sea urchins) are concurrent; if the number of one increases or decreases so does the other (Day and Branch, 2000). Thus the rock lobster is exerting a marked negative impact on abalone by feeding on sea urchins and thereby limiting sheltering habitats for abalone juveniles.

1.3. Abalone farming

1.3.1. Globally

Important abalone fisheries once existed in China, Mexico and New Zealand. Most of these fisheries have now collapsed due to various factors. The only fisheries of any significance that are left are most likely located in Australia and Japan. Due to the collapse of many abalone fisheries, abalone farming has become important as the worldwide demand for abalone is still high. Japan is regarded as the pioneer of abalone farming (Fallu, 1991) and by 1991, abalone culture was being practised in the USA (California), Japan and Taiwan, although the farming methods used in each country vary (Britz, 1991). In 2007, countries with the foremost abalone fisheries were South Africa, New Zealand, Australia, Korea, Taiwan, China and Japan (Fishtech, Inc, online).

Hatchery technology was initially developed after WWII in the hope that reseeding from this supply would boost the fisheries production (Fallu, 1991). Of the currently recognized 56 abalone species found globally, about 20 are harvested commercially (Tang *et al*., 2004). In 2007, the world abalone farm production was 10 000 tons, while the demand was approximately 23 000 tons (Fishtech, Inc, online).

1.3.2. South Africa

Abalone farming is seen as a solution to the threatening abalone crisis in South Africa (Britz, 1991), especially as a ban on all abalone fishing, effective from February 2008, has been instated (Benton, 2007 [online]). South Africa was fortunate as the fundamental techniques for abalone culture had already been established and developed in other countries (Britz, 1991).

Abalone fisheries have been in existence since 1949 spanning 580km of coastline from Quoin Point (East Coast) to Cape Columbia (West Coast) (Dichmont *et al*., 2000). Successful abalone cultivation, however, only started in the 1980s, when it was first demonstrated that *H. midae* could be spawned in captivity (Cook, 1998). The bulk of farms in South Africa have been set up along the south-western part of the coastline. An advantage of these areas is that large kelp beds (main food source) are situated close to the shore. Other advantages include that these areas are relatively free of domestic and industrial pollution and there is reasonably

good water quality. All of the farms in South Africa use a pump-ashore and land-based on-growing system (Cook, 1998).

Of the six species of abalone in South Africa only *H. midae* is harvested and farmed commercially as the others are either too difficult to harvest, or undersized, or their numbers are inadequate and *H. midae* is the largest growing species in size (Cook, 1998). By 2001, thirteen abalone farms had been established in South Africa with an investment of approximately US $12 million (Sales and Britz, 2001) and at present there are fifteen abalone farms (Aucamp, 2007 [online]). In 2006, abalone production in South Africa amounted to approximately 60% of the aquaculture revenues as farm production produced over 900 tons of abalone and the anticipated mark for 2007 is 1000 tons. The abalone aquaculture industry value was estimated to be more than 141 million South African rand ($19.7 million) in 2006, and employed approximately 800 people (Roelf, 2007 [online]).

Farm-produced abalone are not intended for local consumption and are solely produced for the export market. The existing farm permits actually prevent the sale of farmed abalone in South Africa. Abalone products are sold at sizes between 80 and 100mm, which is less than the smallest harvestable size for abalone in their natural environment. The aim of preventing the sale of farm-produced abalone in South Africa is thus to prevent the illegal harvesting of undersize animals (Cook, 1998).

1.3.3. Abalone farming

1.3.3.1. Spawning

Before spawning can occur, the abalone broodstock need to be fully developed internally. Spawning is the event that occurs when the eggs or sperm are released from the ripe adult abalone. Various stimuli can act as triggers; in the wild the stimuli are a change in water temperature, while farms use various techniques to induce spawning, such as UV light, hydrogen peroxide, heat shock, other spawning abalone, and pH shock (Fallu, 1991). Each farm has their personal preference as each technique has its own drawbacks and advantages.

Fertilization is generally completed in about 30 min; the spawning tank during this time looks intensely cloudy due to the high-density of sperm and eggs in the tank. This condition cannot be left for long as it will deplete oxygen levels in the water, which may lead to mortalities of adults and fertilised eggs (Nash, 1991). In farming conditions, static water is used to prevent

loss of eggs or sperm (Fallu, 1991). The fertilized eggs sink to the bottom of the tank where they are easily removed by carefully siphoning them off (Nash, 1991).

1.3.3.2. Larvae

The fertilised egg divides repeatedly and a larva hatches from the egg. The larvae are positively buoyant and float for approximately a day. It is at this stage that the abalone larvae are transferred to a new container containing filtered, sterile and free flowing water, as they are easy to collect by pouring off the top water. They become neutrally buoyant after a day and during this time natural movement distributes them throughout the container as they neither sink nor float (Fallu, 1991).

The larvae do not have any external dietary requirements, as during their planktonic life they feed off their egg reserves. Thus they are free of the problematic dietary requirements that are often linked with larvae of other species (Landau, 1992).

The larvae go through various stages of growth and development, namely the pelagic trochophore, and veliger stages (Fallu, 1991). During the pelagic trochophore stage they can propel themselves using peaks of microscopic cilia. They enter the veliger stage when they start to develop a shell and foot. In the veliger stage a change from a free-swimming organism to one committed to being on a substrate occurs. It is in the veliger stage that, as previously mentioned, they are easily siphoned off into a larger rearing container. This is due to the larvae's positive phototropic state (attracted to light) and therefore they can easily be attracted to a corner of the container by using light (Nash, 1991).

1.3.3.3. Settlement and spat

Before settlement occurs the larvae develop a clearly visible eye-spot. The process occurs approximately five days after fertilisation. The larvae sink to the bottom of the container and take up residence there. The larvae are now known as spat. Once they have settled on the bottom, the spat's body transforms and now looks like a miniature copy of the adult (Fallu, 1991). The transformation process of the spat is known as torsion by which they obtain their permanent shell. Once torsion is completed they are committed to crawling (Nash, 1991).

1.3.3.4. Grow-out

Once the abalone have settled (settlement) they start to feed and grow. Initially they are sexually immature, but the gonads start to develop upon settlement (Fallu, 1991). As the

juveniles grow they become light sensitive and require shelter. It is at this point that they become nocturnal animals (Landau, 1992).

After approximately 3-4 years the adult abalone are collected for export to the overseas market (Nash, 1991).

1.4. Aquaculture genetics

1.4.1. Status

Aquaculture genetics essentially began more than 2000 years ago in China with the start of aquaculture there, coinciding with the Romans who first started to breed fish in ponds. Breeding programmes became more prevalent only when the Japanese started to develop special varieties of koi carp in the 1800s (Dunham *et al*., 2001).

In the 1960s, genetic enhancement programmes were established for the first time, while molecular-based knowledge only surfaced in the 1980s. From the late 1990s there was a marked increase in gene mapping and genomics of aquatic organisms. The genetic response to selection of growth rate and fecundity is superior in aquatic species compared to terrestrial farm animals. These advantages have allowed for improved aquaculture production and higher selection intensity (Dunham *et al*., 2001). Selection and breeding programmes have been successfully implemented for example in Arctic charr, *Salvelinus alpinus* (Ditlecadet *et al*., 2006); oysters, *Crassostrea gigas* and *C. virginica* (Ward *et al*., 2000; Calvo *et al*., 2003; Guo *et al*., 2003); shrimp, *Litopenaeus vannamei* (Gjedrem and Fimland, 1995); tilapia, *Oreochromis niloticus* (Agresti *et al*., 2000) and sea bass, *Dicentrarchus labrax* (Garcia de Leon *et al*., 1998).

1.4.2. Genetic improvement programmes

To date the potential of well-designed genetic improvement programmes leading to significant genetic gains and biological improvements are still to be fully realised. Until recently the focus of aquaculture research was predominantly the environment of the animal; for example, tank design and culture techniques. The focus has now begun to shift towards genetics (Elliott, 2000).

Selective breeding programmes have a vast potential to boost the productivity of aquaculture. These programmes will not only improve the utilisation of available feed, but land and water resources as well, leading to enhanced production (Hayes *et al*., 2006).

Growth is generally the first trait of interest that is targeted. Abalone are slow growers and this becomes a hindrance when considering aquaculture potential. Producing faster growing abalone will reduce production time and cost for market size individuals (Elliott, 2000).

Selective breeding as a means to increase profits for farmers has a good track record, but globally research into aquaculture programmes has been very time-consuming (Lymbery *et al*., 2000; Gjedrem, 2002; Mair, 2002). Only a few studies have reported on growth rate and its heritability in haliotids, but all of these suggested that a selective breeding programme would be successful in increasing growth rate (Hara and Kikuchi, 1992; Kawahara *et al*., 1997; Jonasson *et al*., 1999; Mgaya, 2000; Lucas *et al*., 2006).

1.4.3. DNA markers and their applications

DNA markers, also known as molecular markers, have various applications and contribute significantly to commercial breeding and genetic improvement programmes (Moore *et al*., 1999). Markers consist of variable genes or expanses of DNA (Zeng, 1993; Jansen and Stam, 1994) and can be applied to, for example, the development of genetic linkage maps and characterisation of quantitative trait loci (QTL) (Moore *et al*., 1999).

The majority of performance and production traits are controlled by multiple genes and are thus inherited as quantitative traits (Liu and Cordes, 2004). The loci controlling these traits are universally referred to as quantitative trait loci (Liu, 1998). QTL are not necessarily genes themselves, but stretches of DNA that are linked to the genes that underlie the trait being researched, in this case an economically important trait (Liu and Cordes, 2004). Quantitative traits are important to numerous animal breeders as these traits may influence for example, growth rate and food conversion rate, which are economically important performance indicators (Falconer and Mackay, 1995).

QTL for growth rate have, for example, been identified in *H. rubra* (Baranski *et al*., 2006) and future genome scans should detect QTL for disease resistance in abalone. Growth and disease resistance are the two most common traits selected for QTL analysis, due to their overall significance and importance to aquaculture selective breeding programmes (Hayes *et al*., 2007). One of the most significant applications of linkage and genetic mapping is to find these genes of interest (QTL) and this form of mapping is called QTL mapping (Liu, 1998).

QTL mapping entails the development of genetic maps and identifying an association or relationship between traits of interest and polymorphic DNA markers. If a significant association is found between the markers and traits, it may be an indication of a QTL near the markers (Liu, 1998).

The use of DNA markers in aquaculture research has allowed swift progress in other areas including the assessment of levels of genetic variability and inbreeding, species and strain identification, parentage assignment and marker-assisted selection (MAS) (Liu and Cordes, 2004).

MAS uses DNA or gene markers linked to QTL or major loci in genetic improvement programmes (Davis and Hetzel, 2000; Liu and Cordes, 2004). MAS is the process in which the results obtained from DNA testing are used to assist in selecting individuals with favorable traits to become the parents of the next generation. The selection of animals with suitable genetic variants for a given trait can be done at an early stage, which will improve the rate of genetic progress (van Eenennaam, 2004). MAS will be ideal for traits that are expensive and problematic to measure and also difficult to breed for (i.e. disease resistance and food conversion efficiency) (Davis and Hetzel, 2000). The greatest genetic contribution of MAS in abalone breeding programmes will be for traits such as meat quality and disease resistance that cannot be measured on a breeding candidate (Hayes *et al*., 2007).

To date the application of MAS in aquaculture has been sporadic, as QTL mapping and MAS are not as advanced in aquaculture as in terrestrial animals (Martinez, 2006; Hayes *et al*., 2007). In future the detection of complex traits and the mutations affecting them will greatly increase the value of MAS in fish and shellfish (Martinez, 2006); especially since large full-sibling family sizes (offspring have both parents in common) can be easily obtained in aquaculture species, which is advantageous in breeding programmes (Hayes *et al*., 2007).

Frequently used molecular markers in aquaculture genetics include allozymes, mitochondrial DNA, restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), microsatellite, expressed sequence tags (EST) and single nucleotide polymorphism (SNP) markers (Liu and Cordes, 2004).

1.5. AFLP

1.5.1. General

AFLP is a DNA fingerprinting technique developed by Vos *et al*. (1995). The procedure is based on selective polymerase chain reaction (PCR) amplification of restriction fragments from digested genomic DNA. The restriction sites serve as primer binding sites, thus patterns generated depend on the sequence of the PCR primers and the quality of template DNA used. The AFLP technique does not require any previous sequence knowledge and can be used for DNA derived from any origin or complexity. A limited number of primer sequences may be used and using specific primer sets can increase selectivity (Vos *et al*., 1995).

Fragments obtained by AFLP often correspond to unique positions in the genome and consequently can be exploited in genetic and physical maps by using these fragments as landmarks. Each fragment are characterised by the primers used to amplify the fragment and their size after amplification (Vos *et al*., 1995).

1.5.2. Type II markers

Type I markers are linked to genes of known function, while type II markers are linked to unidentified genomic segments. AFLPs thus represent type II markers as they are usually amplified from genomes of which there is no prior sequence knowledge available. Other markers such as microsatellites and RAPDs fall primarily into the type II marker category as well. Type II markers are considered to be non-coding and consequently selectively neutral. These markers have been used widely in population genetic studies where classifications of genetic diversity and divergence within and between populations are based on supposition of Hardy-Weinberg equilibrium and the selective neutrality of markers used. Type II markers have been shown to be useful for application in aquaculture genetics for species, hybrid and strain detection, in breeding studies and lately as markers linked to QTL (Liu and Cordes, 2004).

1.5.3. Methodology

The general AFLP protocol described by Vos *et al*. (1995) and Mueller and Wolfenbarger (1999) includes the following basic steps: (1) Restriction of the DNA followed by the ligation of oligonucleotide adapters; (2) selective amplification of the restriction fragments; and (3) analysis of amplified fragments using gel electrophoresis.

The fluorescent AFLP molecular technique has been adapted from Vos *et al*. (1995) to facilitate automation and includes the following fundamental steps (Figure 1.5): (1) Template preparation and adapter ligation; (2) preselective amplification; (3) selective amplification; and (4) primer selection for amplification screening (Perkin-Elmer, 1995).

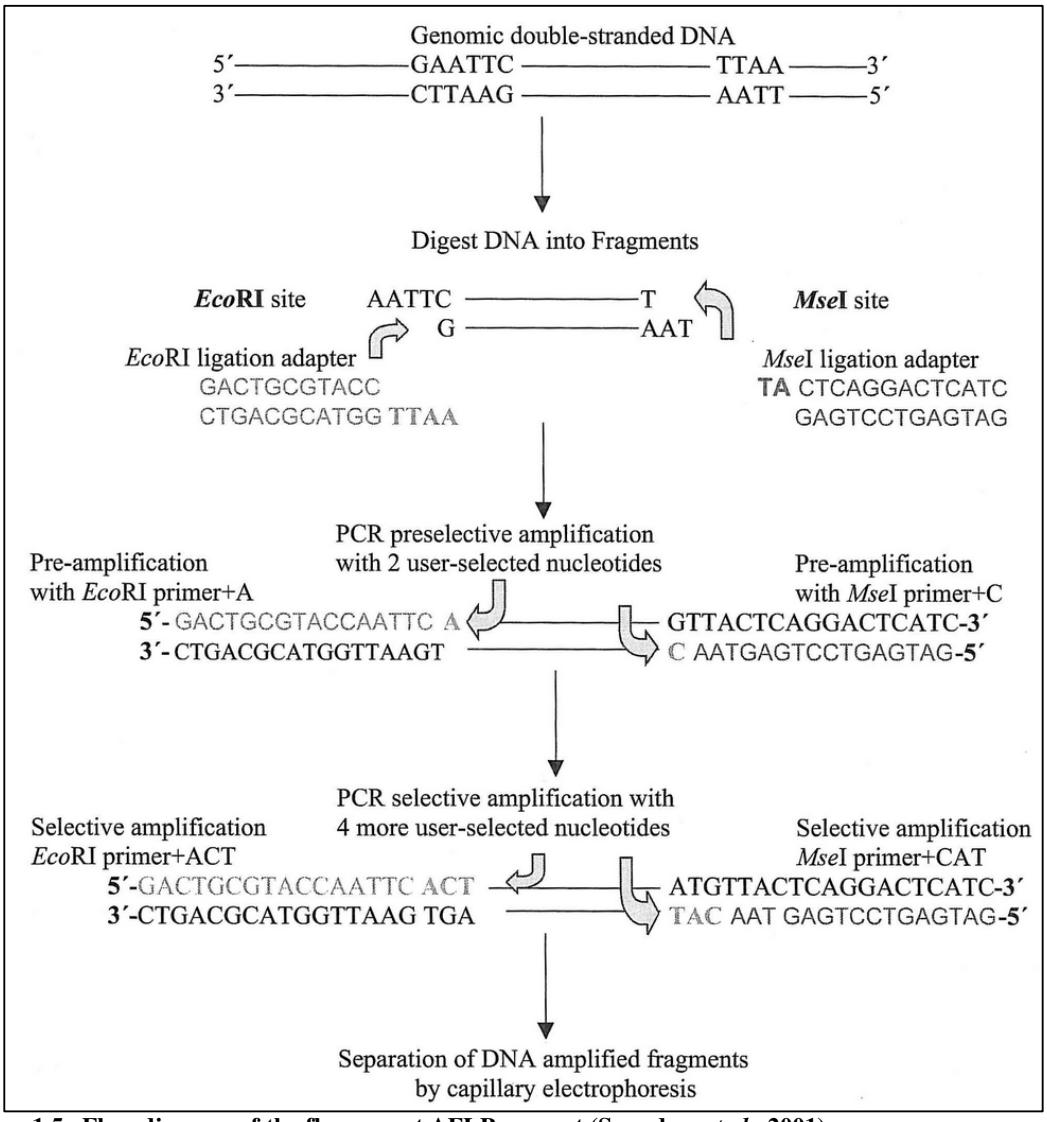


Figure 1.5. Flow diagram of the fluorescent AFLP concept (Saunders *et al*., 2001).

The product obtained is an AFLP genotype, which is a unique, reproducible fingerprint of the individual or organism analysed. The AFLP markers, which produce the fingerprint, are distributed widely throughout the genome allowing for a genome-wide investigation of variation (Meudt and Clarke, 2007).

1.5.4. Advantages

AFLP markers can be produced for any organism, be it animal, plant, human or microbial, making it a potentially universal DNA fingerprinting system. The technique allows for examination of an entire genome and produces genome-wide polymorphic markers without any prior sequence knowledge, thus making it an ideal and powerful technique for generating linkage maps (Mueller and Wolfenbarger, 1999; Liu and Cordes, 2004). Successful AFLP linkage maps have been constructed for molluscs such as oysters (Yu and Guo, 2003; Li and Guo, 2004) and recently in blacklip abalone, *H. rubra* (Baranski *et al.,* 2006) and the Pacific abalone, *H. discus hannai* (Liu *et al*., 2006; Sekino and Hara, 2007).

AFLPs have a low error level as the technique is carried out using especially stringent conditions. Unlike RAPD-PCR where slight differences in the thermocycling parameters may lead to changes in banding patterns, AFLPs are highly reproducible (Mueller and Wolfenbarger, 1999). Small quantities of DNA and partially degraded samples can be used for AFLP. Thus small organisms can be analysed using this technique (Vos *et al*., 1995; Mueller and Wolfenbarger, 1999), which is ideal for the abalone juveniles used in this study. AFLP markers can be identified at a relatively fast rate, which is further facilitated using fluorescent AFLP technology (Perkin-Elmer, 1997; Mueller and Wolfenbarger, 1999). As AFLP markers segregate in a Mendelian fashion, they can be used for QTL analyses and population genetics (Mueller and Wolfenbarger, 1999). Lastly, the nucleotide extensions of the adapters can be altered, thereby identifying hundreds of reliably produced markers (Perkin-Elmer, 1997; Mueller and Wolfenbarger, 1999).

1.5.5. Drawbacks

The main drawback of AFLPs is that it is dominant rather than co-dominant marker type. There are, however, software packages presently available that can help identify co-dominant AFLP peaks. The scoring of co-dominant AFLPs is possible when using well-characterised families, but not for population studies (Liu and Cordes, 2004). The dominant nature of the marker makes it impossible to calculate the departure from Hardy-Weinberg equilibrium (Campbell *et al*., 2003). AFLPs are poorly repeatable between different laboratories and populations and consequently the difficulty in transferring AFLPs significantly decreases the application of AFLP maps generated (Wang *et al*., 2004).

AFLPs are only cost effective for short-term studies. They can be developed quickly and consequently applied sooner than other types of markers, for example microsatellites, which are costly and time consuming to develop (Meudt and Clarke, 2007). However, for long term and extensive projects the cost of AFLPs outweigh their initial ease of development and application.

1.5.6. Scoring and reliability

Fragments generated via AFLP analysis can be visualised and scored using a variety of techniques, namely agarose gel electrophoresis, polyacrylamide gel (PAA) electrophoresis and/or automated genotyping. Agarose gels are very user-friendly with minimum costs, but they provide the least resolution, while PAA manually or together with an automated sequencer provide the best resolution for the detection of single nucleotide length differences (Mueller and Wolfenbarger, 1999). In this study an automated genotyping system will be utilised.

1.5.7. Applications

There are many applications of AFLPs (Vos *et al*., 1995; Perkin-Elmer, 1997; Liu and Cordes, 2004) that include:

- Monitoring the inheritance of traits
- Assessing the degree of variability or relatedness between genotypes
- Diagnostics of genetically inherited diseases
    - Identification of closely-linked DNA markers
    - Genotyp ing of individuals and determination of genetic distance
- Biodiversity studies
- Ascertaining linkage groups in crosses and construction of DNA marker maps
- Saturation of areas of introgression with markers used for gene landing efforts
- Accurate mapping of genes and their subsequent isolation
- Creation of "transcript profiles" for gene expression analysis
- Identification of genomic clones (YACs and BACs) that can be used to construct physical maps

- Forensic analysis
- Parentage and pedigree analysis

1.5.8. Conclusions regarding AFLPs

AFLP markers are superior to markers such as allozymes, RAPDs, RFLPs and microsatellites due to their cost and time efficiency during development, as well as their reproducibility and resolution, although they generate dominant rather than co-dominant markers. AFLPs can generate numerous genome-wide unique polymorphic markers and is a superior technique to RFLPs and RAPDs for generating and saturating linkage maps (Vos *et al*., 1995; Meuller and Wolfenbarger, 1999).

AFLPs are therefore the best tool to reach the goal of this project, which is the development of a preliminary linkage map. The map can be further saturated with co-dominant microsatellite markers.

1.6. Microsatellites

1.6.1. General

Microsatellites are also known as simple sequence repeats (SSRs) as they contain multiple copies of tandemly arranged short sequences, ranging from 1 to 6 base pairs in size (e.g. AC or ACA). They are abundant in all species that have been studied to date (Liu and Cordes, 2004) and have been estimated to occur no less than once every 10kb in fish species (Wright, 1993). The genome of the Japanese pufferfish, *Fugu rubripes*, and related spotted green pufferfish, *Tetraodon nigroviridis*, for example consists of 1.29% and 3.21% microsatellites, respectively (Crollius *et al*., 2000).

Microsatellites are highly polymorphic due to deviation in the number of repeats and tend to be uniformly distributed throughout the genome on all chromosomes and regions of chromosomes. The greater the number of repeats, the more polymorphic the microsatellite is, although as few as five repeats have been observed to be polymorphic. Mutation rates of microsatellites are reported to be as high as 10^{-2}-10^{-6} per locus per generation (Chistiakov *et al.,* 2005) and are thought to be due to polymerase slippage during replication, leading to variation in the number of repeat units (Liu and Cordes, 2004).

Microsatellites frequently represent functionally significant polymorphisms, although they are generally considered to be selectively neutral markers (Liu and Cordes, 2004). Microsatellites are functionally important in: (1) DNA structure, as they are involved in the configuration of numerous DNA structures with simple and complex loop-folding formations; (2) DNA recombination hot spots; (3) influencing DNA replication and lastly; (4) affecting gene expression when located in gene promoter regions (Chistiakov *et al.,* 2005).

Microsatellites, as AFLPs, are mostly type II markers (refer to section 1.5.2) (Liu and Cordes, 2004) as they are generally situated in noncoding intergenic regions (Chistiakov *et al.,* 2005).

In aquaculture, microsatellites are considered to be workhorse markers (Chistiakov *et al.,* 2005). They have been shown to be a powerful tool for precise genetic evaluation of population differentiation and have been used for pedigree assessment in hatchery populations from diverse fishery animals (Li, Q. *et al*., 2006) and parentage determination in aquaculture research (Li *et al*., 2007).

Microsatellites have been successfully developed and characterised for several abalone species, namely *H. asinina* (Selvamani *et al*., 2000), *H. discus discus* (Sekino and Hara, 2001), *H. rubra* (Huang and Hanna, 1998; Evans *et al*., 2000), *H. rufescens* (Kirby *et al*., 1998), *H. kamtschatkana* (Miller *et al*., 2001), *H. discus hannai* (Qi and Akihiro, 2007) and *H. midae* (Bester *et al*., 2004, Slabbert *et al*. [in press]).

1.6.2. Development of microsatellite markers

Identification and characterisation of type II microsatellite markers involve the development of small-insert genomic libraries that are enriched for tandem repeats. The procedure is called an enrichment technique and involves selective hybridisation of disjointed genomic DNA by using a tandem repeat-containing oligonucleotide probe followed by additional PCR extension of the hybridisation products (Chistiakov *et al.,* 2005).

Another procedure that can be followed is to use previously isolated microsatellite markers from species that are closely related. The cross-species amplification allows applying genetic information of one species to another, without having to invest in the detection and characterisation of microsatellites (Chistiakov *et al.,* 2005).

1.6.3. Advantages

Microsatellites have numerous advantages, namely they are relatively abundant, their wide and random genome distribution, high polymorphism and small locus size (Liu and Cordes, 2004; Qi and Akihiro, 2007). These markers have a hyper-variable and co-dominant nature and are easily, rapidly and reliably examined through PCR (Baranski *et al*., 2006; Li, Q. *et al*., 2006; Qi and Akihiro, 2007). The co-dominant property of microsatellites results in a large amount of segregation information and the transferability across populations allows for the establishment of linkage groups among populations and microsatellite maps that offer a high level of portability (Baranski *et al*., 2006; Sekino and Hara, 2007).

Microsatellites' neutrality and the unambiguous scoring of alleles are advantageous (Li, Q. *et al*., 2006). When these markers are studied in a multiplex fashion the turnaround time is relatively fast, with a comparatively low material cost for the user (Li *et al*., 2007). A multiplexed system may be used when products from a range of loci are obtained, which contain non-overlapping arrays of allele sizes and the products can be amplified together efficiently using a standard set of conditions (Chambers and MacAvoy, 2000).

In contrast to dominant markers, microsatellites occur in both coding and noncoding regions and as mentioned, display on the whole a random distribution throughout the genome. Furthermore, they exhibit high levels of both intraspecific and intrapopulation polymorphism (Cristescu *et al*., 2006). All these characteristics promote microsatellite markers as the ideal marker to be used in conjunction with AFLP markers for application in this study.

1.6.4. Drawbacks

The main drawback of microsatellites is the time and effort-consuming process of detecting them, consisting of the isolation and classification of the markers. The detection of microsatellites is more time consuming than the detection of AFLPs (Baranski *et al*., 2006; Qi and Akihiro, 2007). Every microsatellite locus needs to be identified and the flanking region sequenced for the development of PCR primers (Liu and Cordes, 2004). In addition, to successfully isolate microsatellites, information on their abundance and length distribution of repeats in the genome of the organism being studied is required. At present such data is still limited and incomplete for many taxa and in some cases completely absent (Qi and Akihiro, 2007).

Complications with the use of microsatellites for linkage mapping in molluscs have arisen due to frequently occurring null alleles (allele or marker no longer detected as a result of a mutation), segregation distortion and the influence of repetitive elements (Baranski *et al*., 2006).

1.6.5. Scoring and reliability

Amplified microsatellite markers are generally sized and scored on denaturing polyacrylamide gels by using procedures that were initially developed for DNA sequencing. The fragments on the gels may be visualised through ethidium bromide staining, silver staining or autoradiography. Autoradiography procedures detect the ^{32}P or ^{33}P α-labelled nucleotides that have been integrated during the amplification process (Chambers and MacAvoy, 2000).

The use of automated DNA sequencing and fragment sizing together with the related software has become common practice for microsatellite developers, as it results in a faster turnover rate. Fluorescently labelled primers are used that are recognised by software to visualise microsatellite markers for automated scoring (Chambers and MacAvoy, 2000).

Despite the advances, many researches prefer to use manual systems instead of automated systems, as they cannot be positive that the automated software is scoring the data or alleles accurately and consistently. Complications may consequently arise when considering whether or not to combine data obtained using manual versus automated technologies (Chambers and MacAvoy, 2000).

1.6.6. Applications

Microsatellites have been applied to a large variety of research fields and practical disciplines (Liu and Cordes, 2004; Chistiakov *et al.,* 2005). These include:

- Genetic mapping
- Individual DNA identification
- Parentage assignment
- Phylogeny and population genetics
- Conservation genetics

- Molecular epidemiology and pathology
- Quantitative trait loci mapping
- Marker-assisted selection
- Identification of stock structure

1.6.7. Conclusions regarding microsatellites

Microsatellites have become extremely popular markers of choice and are used in a wide range of genetic investigations. In the past decade these markers have become widely used in fisheries research (Liu and Cordes, 2004) and in aquaculture studies they have become the marker of choice for genetic assessments of farmed stocks for application in breeding programmes. Genetically improved stocks may be selected early on, while avoiding inbreeding and increasing positive selection response. Microsatellite loci are effective for construction of genetic frameworks on which other markers and genes may be integrated by using different mapping strategies, namely physical mapping, linkage mapping and genomics tools (Chistiakov *et al.,* 2005).

Microsatellites are more informative and transferable than AFLPs, especially between different laboratories (Yu and Guo, 2005). Therefore, microsatellites are a useful additional marker for this linkage mapping study.

1.7. Linkage mapping

1.7.1. General

Linkage maps document the fundamental order, synteny and spacing of genes or genetic markers on the chromosomes. Linkage maps are developed through the analysis of the co-segregation of allelic forms of various markers within pedigrees; determining how frequently alleles are inherited jointly. These maps therefore infer evidence of the location of genes and markers (Garcia de Leon *et al*., 1998; Davis and Hetzel, 2000; Dekkers, 2004; Chistiakov *et al.,* 2005).

1.7.2. Development

The development and construction of linkage maps entail the placement of co-segregating markers into linkage groups. The distances between linked markers are calculated using the proportion of recombinant genotypes formed (Hartl and Jones, 2001). Genetic recombination is the basis of linkage mapping and is the result of crossovers between homologous chromosomes during meiosis, when segments of the sister chromatids are exchanged between the homologs. Recombination generally occurs randomly on the chromosomes and the likelihood of recombination occurring between different loci is linked with the physical distance between loci. The link between recombination and physical distance is the foundation of linkage mapping (Liu, 1998).

Recombination fraction is the measure used to calculate genetic recombination. For recombination between two loci on a chromosome, it is expected that the greater the physical distance between the loci, the greater the likelihood that they will recombine (Liu, 1998). The recombination limit due to crossover is hypothesised to be 50 percent. If two genes, which are linked, are therefore located further than 50 map units apart, theoretically there is a 100 percent chance of crossover occurring between the two genes on the paired homologs. The crossover between the two homologs will result in the formation of new gene combinations on the chromosome. If the two genes, which are linked, are located less than 50 map units apart, then the likelihood of crossover and recombination occurring is decreased (Klug and Cummings, 2003).

A random copy of each chromosome pair is carried over to the gamete during meiosis. Copies of genes, which are located on different chromosomes, are therefore inherited separately from each other, while those genes that are situated on the same chromosome are not inherited separately, but are passed on jointly or "linked". Genes are therefore tightly linked when they are situated adjacent to each other (less than 50 map units). Genetic distance is generally measured in centiMorgans (cM); 1 cM is taken to be equivalent to 1% recombination between loci (Klug and Cummings, 2003).

1.7.2.1. Genetic map construction

The co-segregation of markers in families are analysed by gene mapping algorithms, which assemble markers into their respective linkage groups, followed by the calculation of the most

probable order of markers within the linkage groups (Hartl and Jones, 2001). The process comprises four steps after data collection (Liu, 1998):

1) Pairwise linkage analysis between all potential two-locus combinations.

The analyses are based on the comparison between the observed and expected frequencies of the genotypic classes under investigation. For a marker to be detectable and useful, it must have a number of different alleles in the mapping population. The potential genotypic classes are identified as a function of the number of alleles present at the two loci under consideration and of the mating design, which was used to obtain the mapping population. Recombination fractions are generally estimated using the observed genotypic frequencies and applying a maximum likelihood approach.

2) Allocating the markers to different linkage groups.

The calculated recombination fraction, significance level of the recombination fraction and available genome information, i.e. number of chromosomes, are the criteria used for generating linkage groups.

The number of linkage groups should be similar to the haploid chromosome number of the organism under investigation. The number of linkage groups obtained relative to the haploid chromosome number is dependent on the number of markers used (level of genome coverage achieved), how informative the data is, the genetic model used for data analysis and the grouping criteria.

3) Ordering of markers in the same linkage group.

The relative location of markers on the genetic map is determined. The ordering of markers is the most demanding computational step.

4) Estimation of the multipoint recombination fractions between neighbouring loci.

The estimated multipoint recombination fraction may be slightly different to the two-point recombination fraction and matches the physical distance between loci more closely. Generally, due to the sequence specificity of recombination events, the genetic and physical distance do not have a one-to-one relationship.

1.7.3. In Aquaculture

Through the past decade the number of genetic linkage maps developed have increased for many organisms, especially for many aquatic species including, rainbow trout, *Oncorhynchus*

mykiss (Young *et al*., 1998; Sakamoto *et al*., 2000), black tiger shrimp, *Penaeus monodon* (Moore *et al*., 1999; Wilson *et al*., 2002), yellowtails, *Seriola quinqueradiata* and *S. lalandi* (Ohara *et al*., 2005) and tilapia, *Oreochromis niloticus* (Kocher *et al*., 1998; Agresti *et al*., 2000), *O. aureus* and *O. mossambicus* (Agresti *et al*., 2000).

Progress has been relatively slow in this area with regard to molluscs. Preliminary linkage maps have only been developed recently for the Pacific oyster, *Crassostrea gigas* (Hedgecock *et al*., 2002; Li and Guo, 2004), the eastern oyster, *C. virginic*a (Yu and Guo, 2003) and the Zhikong scallop, *Chlamys farreri* (Wang *et al*., 2004, 2005). For abalone, linkage maps have been developed for only a few species, namely blacklip abalone, *H. rubra* (Baranski *et al*., 2006; based on microsatellite markers) and the Pacific abalone, *H. discus hannai* (Sekino and Hara, 2007; microsatellite markers) (Liu *et al*., 2006; largely AFLP markers).

The slow progress of genetic improvement programmes and linkage map development in molluscs have largely been due to a lack of genetic resources, namely a shortage of molecular markers, and not a lack of interest. There is a large amount of commercial interest in marine molluscs, such as scallops, oysters and abalone, as these molluscs form part of and support vast aquaculture industries worldwide, and are consequently important and lucrative commercial commodities (Li and Guo, 2004).

A preliminary genetic map is an essential requirement for detailed genetic studies of any organism. AFLPs can be used to develop a reasonably dense linkage map rapidly. These markers are, however not always conveyable to new populations and even to crosses in the same population (Li *et al*., 2003; Hubert and Hedgecock, 2004). Microsatellites, on the other hand, negate this problem and are more easily portable (Love *et al*., 1990; Hubert and Hedgecock, 2004). These markers thus form a superior combination for linkage mapping.

1.7.4. Applications

Uses of linkage maps (Wilson *et al*., 2002; Wang *et al*., 2004; Ning *et al*., 2007) include:

- The study of genome structure and evolution
- Comparative genome mapping
- Determining the genetic basis of heterosis

- Identification and mapping of quantitative trait loci
- Marker-assisted selection
- Positional cloning of genes

1.8. Statistical analysis

1.8.1. Linkage mapping

1.8.1.1. Computer software

Computer technology and the advances that the technology has undergone form an integral part of the development of genetic maps and are largely responsible for the considerable progress in this field. Theories originally developed for classical mapping can be applied to mapping of DNA markers, as the concepts remain the same, i.e. genetic distance is measured in centiMorgans and is based on the occurrence of genetic crossovers. The difference between classical and molecular-based mapping is the number of markers used, as the number of markers that can be mapped in a single population during DNA-based mapping can without difficulty reach into the thousands, consequently leading to the evolution in computer technology to handle these large datasets (Young, 2000).

Basically, to develop a linkage map from DNA marker data, statistics software packages are needed, which are able to run chi-squared contingency table analyse efficiently. Two-point (pairwise) linkage between markers forms the foundation for constructing linkage maps and is calculated by the chi-squared statistical test (Young, 2000). Two-point statistics only take two markers into consideration at a time (Cartwright, 2007) and are generally used to group markers into linkage groups, i.e. identifying linkage between two loci. Genetic mapping algorithms subsequently reconstruct and determine the order of the markers and the genetic distances (in centiMorgans) between these markers on the chromosomes as precisely as possible (Cartwright *et al*., 2007).

There are many statistic software packages available and they fall into two categories: (1) genetic mapping algorithms that use multipoint likelihood maximisation, i.e. MapMaker (Lander *et al*., 1987) and CRI-MAP (Green *et al*., 1990); (2) the other category uses two-point statistics, i.e. Gmendel (Echt *et al*., 1992) and JoinMap (Stam, 1993). Multipoint analysis has advantages over two-point analysis, as it can construct a linkage map in the presence of missing data of an individual for any of the markers, while two-point statistics cannot.

Multipoint analysis examines numerous loci, unlike two-point analysis, which only analyses two loci at a time. Thus, multipoint analysis is more advantageous when using markers that are not fully informative. However, two-point analysis is a faster technique than multipoint analysis (Cartwright *et al*., 2007). As a result, two-point analysis is commonly used to generate linkage groups, while multipoint analysis is used to determine the order of the markers in the linkage groups, as multipoint analysis can analyse all the different marker orders possible and identify the most likely marker order (Lander *et al*., 1987; Young, 2000). For example, if there are three different markers, the multipoint analysis will compare all possible combinations of the three markers (for three markers there are six different possible combinations) and compute the best order. In Mapmaker 3.0, the log-likelihood of the other orders is compared to the best order found (maximum likelihood maps for each possible order of markers is computed and arranged by the likelihood of their maps), which should have a relative log-likelihood of 0.0 and the less likely map orders will have a negative relative log-likelihood for example –0.8 (Lincoln *et al*., 1992).

Constructing a linkage map using DNA markers is usually the initial step in genetic marker analysis. Ideally one will be able to relate the map constructed to those obtained from other mapping populations and for QTL analysis (Young, 2000).

1.8.2. Principles and statistics used in linkage mapping

1.8.2.1. Mendelian genetics

Mendelian or transmission genetics forms an integral part of linkage genetics and simple trait segregation theory. Genes, which are units of heredity, are passed along from generation to generation in a sexually reproducing species, following Mendelian inheritance (Liu, 1998).

In this study the segregation of markers (the segregation of the parent alleles through the offspring) will be analysed according to expected Mendelian segregation ratios.

1.8.2.2. Chi-square analysis

For genomic mapping there are numerous statistical analyses that are commonly used, namely Poisson, binomial, normal and chi-square (χ^2) distribution (Liu, 1998). For this study, chi-square analysis will be discussed further.

Chi-square analysis is widely used in genomics and measures goodness of fit of the null hypothesis and evaluates observed deviations. The null hypothesis postulates that there is no

significant difference between the measured values and the expected values. Essentially, any apparent difference between measured and expected values can be accredited to chance. If the null hypothesis is rejected then any difference between the two values cannot be attributed to chance (Klug and Cummings, 2003).

The chi-square test reduces to a single numerical value the observed deviation in each part of an expected ratio and takes into account the sample size. The larger the sample size, the smaller the impact of chance deviation on the outcome. The χ^2 value is used to determine how much of the observed deviation can be attributed to chance only. The formula for chi-square analysis is:

$$\chi^2 = \Sigma \quad \frac{(o-e)^2}{e}$$

Where:

o = the observed amounts for a particular category

e = the expected amounts for the specified category

Σ = the sum of the calculated amounts in the ratio

The final step in the analysis is to determine the degrees of freedom (df) and interpret the χ^2 value. The degrees of freedom is calculated using the number of different categories in the dataset being analysed and is important as the greater the number of different categories, the greater the deviation expected due to chance. Degrees of freedom equals n-1, where n is the number of categories. Once the degrees of freedom have been calculated, the χ^2 value can be interpreted in terms of a probability value (P), which is generally determined using a graph or table (Klug and Cummings, 2003).

The significance level or P value is generally set at 0.05-0.01 (Liu, 1998). For example a P value less than 0.01 indicates that the likelihood is less than 1% that the observed deviation in the results could be acquired by chance alone (Klug and Cummings, 2003) and therefore any deviations between the observed and expected values are significant.

1.8.3.3. Pearson correelation

The Pearson product moment correlation coefficient, *r*, [designated with an "*r*" when computed in a sample and "*p*" (Greek letter rho) when measured in a population] is a dimensionless index that ranges from –1.0 to 1.0. A correlation coefficient displays the degree to which two datasets or variables (*X* and *Y*) are related (closeness of the relationship) and reflects the degree of linear relationship between the two variables under investigation. A correlation of +1 indicates a perfect positive relationship (correlation) between the two datasets or variables (the two variables increase together), while a correlation of –1 means there is a negative relationship between the two variables. An *r* value of 0 indicates that there is no correlation between the two variables. The number of paired observations used in the calculation of r is used to derive an associated probability statistic (Snedecor and Cochran, 1989).

1.8.3.4. Bonferroni correction

The Bonferroni correction is a mathematical multiple-correction used to reduce falsely significant results in statistical analyses when carrying out multiple tests simultaneously (Weisstein, [online]) and is suitable for sample sizes ranging from 25 to 200 (Sokal and Rohlf, 1995). Multiple tests of statistical significance will give rise to 1 out of every 20 hypothesis-tests appearing significant at the $P = 0.05$ level, simply as a result of chance (Weisstein, [online]).

A given *P* value (probability value) may be suited to individual comparison; it is however not ideal for all comparisons. To avoid obtaining a number of spurious positives, the alpha value is decreased to account for the number of comparisons conducted. To adjust the *P* value threshold when carrying out multiple tests to a stricter threshold, the *P* value is divided by the number of independent hypothesis tests (*n*) being conducted (Weisstein, [online]). For example, if 10 independent tests were being carried out, instead of a *P* value threshold of 0.05, one would use the threshold of 0.005 for each test.

One of the drawbacks of applying the Bonferroni correction is, however, that it increases the type II error rates. A type II error occurs when the null hypothesis is accepted when it is in fact false, thereby creating a false negative (Liu, 1998). To compensate for the increased likelihood of producing type II errors when applying Bonferroni correction to a large number

of tests, such as during linkage mapping analysis, the correction is based upon the number of linkage groups (haploid chromosome number expected [Woram *et al*., 2004]).

Bonferroni correction has been widely used in linkage map development in aquaculture, for example, in blacklip abalone, *H. rubra* (Baranski *et al*., 2006); in pacific oyster, *Crassostrea gigas* (Hubert and Hedgecock, 2004); in Artic charr, *Salvelinus alpinus* (Woram *et al*., 2004); in a comparative analysis of the rainbow trout (*Oncorhynchus mykiss*) genome with two other species of fish, Artic charr, *S. alpinus*, and Atlantic salmon, *Salmo salar* (Danzmann *et al*., 2005) and in Pacific abalone, *H. discus hannai* (Liu *et al*., 2006).

1.8.3.5. Segregation distortion

The definition of segregation distortion loci (SDL) is chromosomal regions that cause distorted segregation ratios. SDL are generally identified by linked markers that show non-Mendelian segregation (Zhu *et al*., 2007).

Segregation distortion is a problem that is often encountered in mapping populations and occurs when markers do not segregate according to the expected Mendelian segregation ratio, i.e. 1:1 or 3:1. It is detected by doing chi-square analysis to compare all the observed progeny ratios against the expected ratios (Baranski *et al*., 2006; Shen *et al*., 2007).

The two primary causes of distorted segregation ratios are null alleles and zygotic viability selection (Reece *et al*., 2004). Non-Mendelian segregation ratios can have technical i.e. genotyping errors, or biological causes such as genetic isolation in parental populations and inbreeding depression (Bratteler *et al*., 2006). Segregation distortion has been frequently related to sterility genes, gamete genes (reduce gamete viability) and chromosome translocations. The degree of segregation distortion displayed by molecular markers linked to the SDL depends on the position and size of the SDL and consequently it is possible to locate the SDL using the distortion (Zhu and Zhang, 2007).

Segregation distortion has been widely reported in bivalve molluscs (Yu and Guo, 2003; Li and Guo, 2004) and in the few *Haliotis* maps developed to date; *H. discus hannai* (Liu *et al*., 2006; Sekino and Hara, 2007) and *H. rubra* (before Bonferroni correction) (Baranski *et al*., 2006).

1.8.3. Genome coverage

Fundamental characteristics of an organism are its genome length (in Morgans or centiMorgans) and its diploid chromosome number. The genome length was classically estimated using chiasma counts in meioses (Chakravarti *et al*., 1991).

The genetic length (Morgans; *G*) can easily be calculated if a dense linkage map of each chromosome is available and is calculated as the sum of all the mapped intervals. However, a preliminary estimate of *G* is useful when regions of the genome are not covered due to a low marker number (partial or incomplete genetic maps) (Chakravarti *et al*., 1991). Hulbert *et al*. (1988) proposed a simple and useful method for estimating *G* when only partial linkage data is available.

The method employed by Hulbert *et al*. (1988) is a method-of-moments type estimator of *G*, where *G* is determined by calculating the observed and expected number of locus pairs, which are greater than a specified log og the odds (LOD) score, i.e. LOD 3. An advantage of the method is that no prior knowledge of the chromosome number is required. Drawbacks, however, are that the properties of the estimation procedure are unknown and the standard error of the estimate cannot be easily attained (Chakravarti *et al*., 1991).

Another estimate of *G* suggested by Chakravarti *et al*. (1991) uses a maximum likelihood method and assumes that there is no interference and the chromosomes are of equal length. These assumptions facilitate the calculation of the variance and confidence limits of *G*. A disadvantage of the maximum likelihood method is that knowledge of the chromosome number is required, although there is a possibility of estimating the chromosome number (Chakravarti *et al*., 1991).

The maximum likelihood method is superior to the method-of-moments when the data are incomplete or when the chromosome lengths differ. When the linkage data are relatively comprehensive, both procedures, however, perform equally well (Chakravarti *et al*., 1991).

Fishman *et al*. (2001) determined the estimated genome length by first calculating the average framework marker spacing (*s*), which is determined by dividing the total length of all the linkage groups by the number of intervals present (number of markers minus number of linkage groups). The value of *s* is used to account for chromosome ends past the terminal

markers, and 2*s* therefore added to the length of each linkage group. The total of all the newly determined linkage group lengths (with 2*s* added) is the estimated genome length.

In this study several techniques were used to estimate genome length: the method described by Fishman *et al*. (2001) and method four as described by Chakravarti *et al*. (1991).

1.9. Objectives

The aim of this study is to identify AFLP-based markers, to monitor their segregation in a single full-sib family and to use these markers and additional microsatellite markers to generate a first genetic linkage map for *H. midae*. The aim will be achieved through the use of fluorescent AFLP technology and automated microsatellite genotyping. Optimisation of the AFLP technology will be carried out by testing for optimal fluorescent primer combinations. The microsatellite markers have been identified and developed by Slabbert *et al*. (in press). Applicable software will be utilised to generate the preliminary linkage map using the data obtained from the fluorescent AFLP-based and microsatellite markers.

The outcome of this study will provide information that can be incorporated in the abalone enhancement programme. Upon its completion the preliminary linkage map upon may be used to locate markers that are linked to the QTL of interest.

Future objectives of the abalone enhancement programme, of which this study forms a small part, are the saturation of the preliminary linkage map with a large number of additional markers as well as different types of markers, for example SNPs that have been developed in separate studies and would provide a denser and more informative map.

CHAPTER 2

DNA EXTRACTIONS

CHAPTER 2 - DNA EXTRACTIONS

The work described in this chapter has been published as a short communication in the Journal of Aquaculture Research (Badenhorst and Roodt-Wilding, 2007; Appendix A).

Construction of a linkage map requires first and foremost the isolation of suitable DNA starting material, regardless of the kind of population or DNA markers being used. There are numerous DNA extraction methodologies available, but care must be taken to isolate DNA of suitable quantity and quality, while aiming for simplicity and speed. The use of AFLP technology, as in this study, is especially troublesome as high molecular weight DNA is required (~100-1000ng) and the DNA must be free of contaminants, such as non-target DNA or/and inhibitory compounds. The contaminants may cause problems in any of the steps (digestion, ligation and amplification) of the AFLP analysis and can result in reduced gel resolution or artifactual peaks (Young, 2000; Meudt and Clarke, 2007).

DNA suitable for molecular studies of abalone is commonly obtained from the epipodia (sensory tentacles) using non-destructive sampling techniques (Slabbert and Roodt-Wilding, 2006). These can be sampled from animals older than approximately one year (± 3-4cm). In studies that have time constraints, alternative sources of DNA must be considered. Abalone larvae (1-5 days) and juveniles (± 5 days-1 year) are considered here as possible DNA sources for fluorescent AFLP analysis. As small amounts of tissue are involved and high-quality DNA is required for AFLPs, variations of the CTAB extraction protocol (usually used for the adult epipodia) (Li and Guo, 2004) and lysis (extraction) buffer protocols (Walsh *et al*., 1991; Simpson *et al*., 1999; Gruenthal and Burton, 2005; Ki *et al*., 2005; Vadopalas *et al*., 2006) are considered.

2.1. Materials and Methods

2.1.1. Sampling

A single full-sib family (F_1) was used in this study. The mature female and male parents were chosen at random from hatchery broodstock (the broodstock are obtained from wild abalone populations). The family was bred and reared at the I&J abalone hatchery (Gansbaai, South Africa). Parental material of the reference family was obtained by non-destructive sampling (Slabbert and Roodt-Wilding, 2006) of 2-3 epipodia (sensory tentacle) per parent. Progeny

(F_1) of the reference family were randomly sampled at 5 days old (veliger larvae), 2- (± 2mm) and 3.5- (± 5mm) months-old (juveniles). All sample materials were stored in 100% (v/v) ethanol at 2-6°C. Individual larvae were sampled under a stereomicroscope, using a micropipette.

2.1.2. DNA extraction

(A) DNA of adult and juvenile (2- and 3.5-month-old) abalone were extracted using a variation of the protocol described by Li and Guo (2004) (based on Saghai-Maroof *et al*., 1984). Whole tissue of the progeny was used where the tissue could be removed from their shells, while a single epipodia was used for the parental (mature) tissue. DNA was extracted from these tissues (epipodia and whole tissue) using the phenol-chloroform protocol (Li and Guo, 2004); the tissue was placed in 700µl lysis buffer [2% (m/v) CTAB, 100mM Tris-HCl (pH 6.8), 1.4M NaCl, 20mM EDTA (pH 8), 0.2% (v/v) 2-ß mercaptoethanol] and 5µl 10mg/ml proteinase K, briefly vortexed and incubated at 60°C overnight. DNA was extracted once with phenol-choloroform-isoamyl alcohol (25:24:1), twice with cholorofom-isoamyl alcohol (24:1) and then precipitated with cold isopropanol overnight in a freezer. Pellets were washed twice in 70% (v/v) ethanol, dried, and suspended in 50µl sterile ddH_2O and stored at 2-6°C.

As only microscopic quantities of tissue are available from larvae, the CTAB extraction protocol usually used for adult abalone (epipodial extraction; DNA extraction method A above) was not appropriate.

Alternative protocols were tested for the larvae:

(B) The Chelex protocol as described by Walsh *et al*. (1991): A single larva together with 200µl extraction buffer [10% (v/v) Chelex 100 resin solution, 10mg/ml proteinase K] was placed on an orbital incubator at 56°C for 30 min in a 1.5ml eppendorf tube, followed by a 95°C incubation for 10 min in a heat block. The supernatant was collected and stored at -20°C.

(C) A variation of Ki *et al*. (2005), using 30µl extraction buffer (autoclaved distilled water, 10mg/ml proteinase K) and a single larva in a 0.5ml micro-centrifuge tube. A one-step incubation/denaturation protocol was used, incubating the sample for 50 min at 55°C followed by denaturation at 95°C for 10 min. The extract solution was stored at 2-6°C.

(D) The protocol as described by Vadopalas *et al*. (2006), where single whole larva were incubated in a 200µl lysis buffer [10mM Tris-HCl (pH 8.3), 50mM KCl, 0.5% (v/v) Tween 20] and 10mg/ml proteinase K, for 2 h at 55°C followed by 30 min at 95°C.

In addition to these methods, protocols as described by Simpson *et al*. (1999) (E) and Gruenthal and Burton (2005) (F) which uses an identical lysis buffer as described in method D, were tested. The only difference from the previous extraction protocols was the extraction volume and a different one-step incubation/denaturation time: 15µl extraction buffer, 60 min at 65°C followed by 15 min at 94°C (E) and 10µl lysis buffer, 1 h at 60°C followed by 20 min at 80°C (F).

(G) Three protocols (methods D, E and F) were later modified to include 50mM EDTA in the lysis buffer as EDTA inhibits the action of DNAses, which break down genomic DNA.

(H) Products of the extractions using a lysis buffer (methods D, E and F) were, in addition, also precipitated with 100% ethanol after the incubation/denaturation step. The pellets were washed twice in 70% ethanol, dried, and suspended in 50µl sterile ddH_20 to determine if more concentrated genomic DNA was obtained with the additional precipitation step compared with using only the lysis buffer.

(I) A variation of the protocol described by Mo and Rinkevich (2001) was used, after isolation of larval DNA using the lysis buffer (from methods D-F). After completion of the incubation/denaturation step (60 min at 65°C followed by 15 min at 94°C), DNA was extracted once with equilibrated phenol (pH 8), once with phenol-chloroform-isoamyl alcohol and finally, a chloroform-isoamyl alcohol step. The DNA was precipitated with 100% ethanol, the pellet washed twice with 70% ethanol and the extracted DNA was dissolved with 40µl sterile distilled water.

All DNA concentrations (wavelength 260/280) and purities (wavelength 230/260) were measured using a NanoDrop spectrophotometer (ND-1000 Spectrophotometer; NanoDrop Technologies) and when concentrations were > 100ng/µl they were adjusted to 100ng/µl for AFLP analysis. A DNA sample was considered to be of high-quality when the 230/260 wavelength was > 260/280 wavelength, and the value of the 260/280 wavelength was ≥ 1.8. A further criterion was a significantly different graph trough and peak of the DNA sample being measured.

2.2. Results

With the older juveniles (method A), 70-100ng/µl and with the 2- and 3.5-month-old juveniles, > 300ng/µl DNA yield was obtained. The 3.5-month-old juvenile samples yielded the highest concentration of DNA and were adjusted to 100ng/µl for AFLP analysis (Appendix B). Furthermore, the majority of the 3.5-month-old samples yielded uncontaminated DNA, which contained no inhibitory compounds or proteins. These extracts could be visualised through electrophoresis on an agarose gel stained with ethidium bromide (Figure 2.1) unlike the larvae extracts, which could not be visualised at all on an agarose gel.

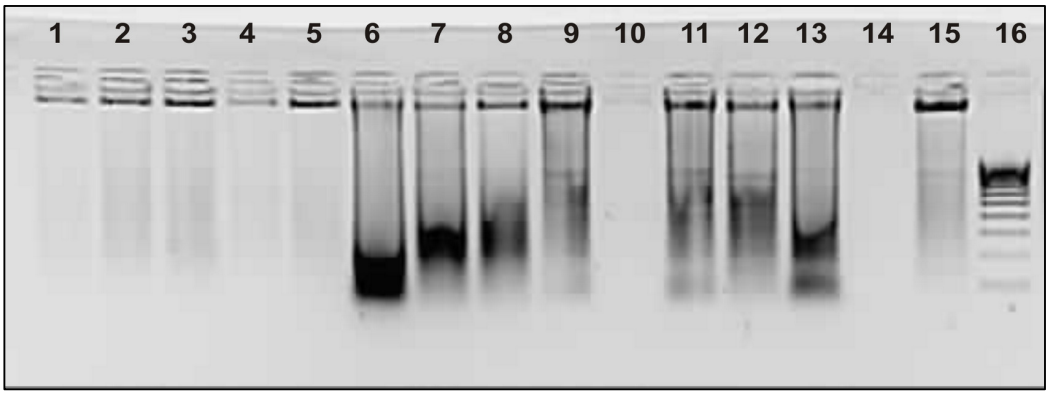


Figure 2.1. Agarose gel depicting the quality of the juvenile genomic DNA extractions obtained by using the protocol of Li and Guo (2004). Lanes 1-5 contain extracted DNA from 2-month-old juvenile samples, while lanes 6-15 contain the 3.5-month-old juveniles' extracted DNA samples. Lane 16 contains a 1kb DNA Ladder (Bioline, London, UK).

For the larvae, the Chelex extraction (method B); the distilled water protocol (method C); the additional precipitation steps (method H) and the added phenol-chloroform-isoamyl wash steps (method I) yielded the poorest results (< 8ng/µl) and were subsequently not further considered for AFLP analysis. The Simpson *et al*. (1999) and Gruenthal and Burton (2005) protocols with added/additional 50mM EDTA (method G) yielded superior results (± 50-80ng/µl) compared with the protocols using only lysis buffer without EDTA (± 20-40ng/µl) (methods E and F) and the Vadopalas *et al*. (2006) protocol (± 40ng/µl) (method D). The superior extracts (method G) and those from methods D, E and F were subsequently used for fluorescent AFLP analysis (Table 2.1).

Table 2.1. Summary of samples and DNA extraction protocols used and results of the extractions and fluorescent AFLP analysis.

Samples	Protocol	Extraction type	Method	DNA extraction results	AFLP analysis	AFLP analysis results
2-month-old juveniles (± 2mm)	Li and Guo, 2004	CTAB extraction	A	Good (± 70-100ng/µl)	Yes	Poor results
3.5-month-old juveniles (± 5mm)	Li and Guo, 2004	CTAB extraction	A	Best results (> 300ng/µl)	Yes	Good
Larvae (5 days old)	Walsh *et al.,* 1991	Chelex	B	Poor (< 8ng/µl)	No	N/A
	Ki *et al*., 2005	A lysis buffer	C			
	Vadopalas *et al*., 2006		D	Average (± 40ng/µl)	Yes	Poor (No visible peaks)
	Simpson *et al.,* 1999		E	Average (± 20-40ng/µl)	Yes	
	Gruenthal and Burton, 2005		F			
	Methods C, D & E	A lysis buffer + added EDTA	G	Good (± 50-80ng/µl)	Yes	
	Methods C, D & E	A lysis buffer + additional precipitation steps	H	Poor (< 8ng/µl)	No	N/A
	Mo and Rinkevich, 2001	A lysis buffer + wash steps	I			

2.3. Discussion

2.3.1. Juvenile abalone

As larvae were not yielding workable extracts for AFLP analysis, juvenile abalone were considered. Two age groups were tested; 2-month and 3.5-month-old juveniles. These two age groups were considered because, unlike the larvae, the juvenile abalone are easily visible to the naked eye at these stages (although still quite small; ± 2-5mm) and since they are much larger than the microscopic abalone larvae, it was thought that greater quantities of genomic DNA and better quality DNA could be obtained compared to the larvae.

The 2-month-old juveniles (± 2mm) were extracted whole with their shell as they were too small to remove from the shell, while the 3.5-month-old juveniles (± 5mm; Figure 2.2 and 2.3) were large enough for the shell to be removed easily. The extracts obtained were of a higher quality and purity than those obtained from the larvae, although the 3.5-month-old

juveniles yielded the best results. These extracts were tested for AFLP analysis and it was found that the 3.5-month-old juveniles yielded good results with AFLP analysis, while the 2-month-old juveniles yielded no results. As the 2-month-old juveniles were extracted with their shells this may have resulted in inhibiting components being present and hindering the AFLP procedure.

Figure 2.2. Size of 3.5-month-old abalone compared to a matchstick.

Figure 2.3. Size of a 3.5-month-old abalone (4.64mm) measured using a calliper.

2.3.2. Abalone larvae

In abalone, DNA is commonly obtained from the epipodia (sensory tentacles). Abalone require approximately a year of growth before they reach a size at which the epipodia can be sampled without difficulty. Methods were therefore investigated in which samples could be taken at an earlier stage so time, waiting for them to grow to an appropriate size for epipodial sampling, would not be an issue. Various sizes, from larvae to juveniles younger than one year, were therefore considered for extraction.

Abalone larvae undergo settlement five days after hatching from their eggs; consequently larvae can be sampled between 1-5 days old. In addition, according to Vos *et al*. (1995), small quantities of DNA, i.e. 2.5pg, can be used effectively for AFLP analysis and low DNA concentrations extracted from larvae were sufficient for PCR and microsatellite analysis in the study by Vadopalas *et al*. (2006). Scallop larvae that yielded concentrations of approximately 5ng/µl were successfully used in an AFLP study by Wang *et al*. (2004). Thus larvae were considered a viable option for this study.

Veliger larvae at five days old were chosen, as the younger larvae were harder to differentiate individually under the stereomicroscope. The 5-day-old larvae had already developed shells and were found to yield more DNA than their younger counterparts.

Extractions (Methods B-I) carried out using the 5-day-old larvae yielded low concentrations and poor quality DNA. Although AFLP analysis can be carried out using low concentrations of DNA, it is important that the DNA is uncontaminated (no inhibitory compounds or proteins) and of good quality. Various extraction methodologies were therefore tested, as described, to determine if results could be improved. Additional steps were added to the extraction methodologies found in the literature, such as EDTA (method G) as EDTA has been noted to prevent DNA degradation in lysis buffers containing TRIS-HCl, KCl and Tween and EDTA inhibits the action of DNAses in general.

In method (H) additional precipitation steps were added in an attempt to remove the lysis buffers' reagents and yield a purified product as the components of the lysis buffer may interfere with the AFLP analysis at the restriction-ligation steps.

In method (I) the same principle was followed as above, by adding additional phenol/cholorform steps in an attempt to purify the extract and to remove proteins still in solution.

None of the methodologies tested on the 5-day-old larvae (methods B-I) yielded DNA extracts suitable for AFLP analysis. This is contrary to studies with scallop where larval DNA was used successfully for AFLP analysis (Wang *et al*., 2004). Scallop larvae are, however, morphologically different from abalone larvae and Wang *et al*. (2004) also used the AFLP protocol as described by Vos *et al*. (1995). The AFLP plant mapping kit used in this study may be more sensitive to contaminating components and poor quality DNA than the AFLP procedure described by Vos *et al*. (1995).

Even though the extracts obtained from larvae using the mentioned lysis buffers are sufficient for PCR (Vadopalas *et al*., 2006), it appears to be inappropriate for fluorescent AFLP analysis. This may be due to the extracts being poor quality (partly degraded) DNA resulting in the restriction-ligation reaction being impeded. The DNA obtained from larvae using these lysis buffers may also not be enough for fluorescent AFLP analysis, as ideally 100ng/µl is required. This quantity could probably be less for abalone tissue as a fluorescent AFLP trial run carried out using DNA obtained from adult abalone epipodia, adjusted to 20ng/µl, gave good results.

It is concluded that the CTAB extraction protocol (method A) yielded the best results with parent and juvenile abalone and that the smallest size and youngest age for sampling juvenile abalone for fluorescent AFLP applications is ± 5mm and 3.5-month-old, respectively.

CHAPTER 3

AFLP MARKER DEVELOPMENT & ANALYSIS

CHAPTER 3 - AFLP MARKER DEVELOPMENT & ANALYSIS

AFLP analysis has come a long way since its introduction in 1995 and has become the preferred marker system for numerous studies on plants and lately for animals, fungi and bacteria. The AFLP markers are produced through restriction endonuclease digestion of genomic DNA, followed by a selective amplification step (Figure 3.1) and electrophoresis of a subset of fragments, thereby generating a DNA fingerprint. The DNA fingerprint is unique for each individual and is reproducible. AFLP markers are dominant markers and are relatively fast and easily produced, informative (due to their sheer number) and require no prior sequence knowledge, making them ideal for genomic studies (Meudt and Clarke, 2007).

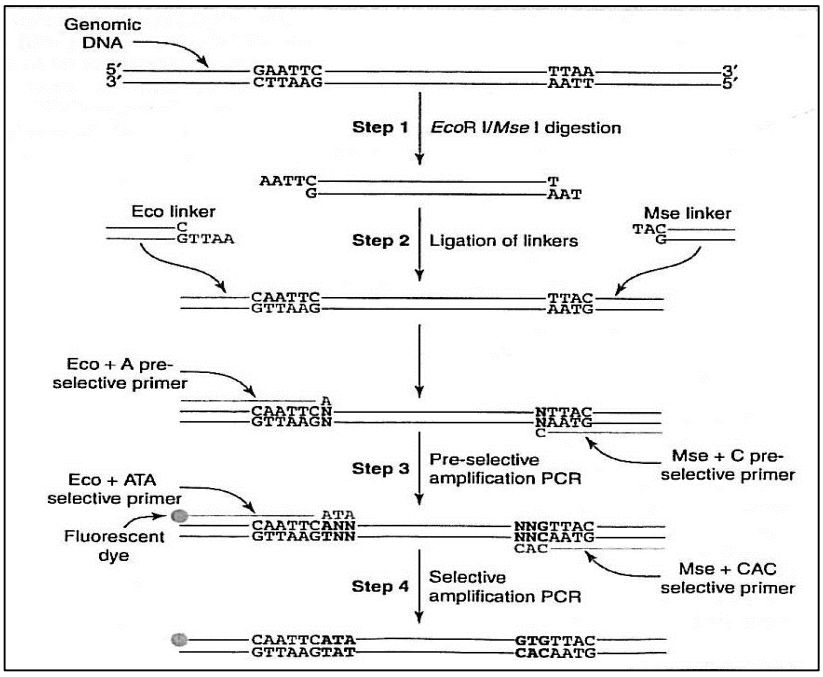


Figure 3.1. Overview of the AFLP procedure (Meudt and Clarke, 2007).

Automated genotyping is slowly replacing gel-based systems, due to their high throughput and superior data quality. Fluorescent AFLP analysis allows for rapid data analysis and generation of numerous AFLP markers, ideal for linkage mapping (Meudt and Clarke, 2007).

3.1. Materials and Methods

The AFLP protocol originally described by Vos *et al.* (1995) was conducted according to Perkin-Elmer's AFLP plant mapping kit protocol (Applied Biosystems) with some

modifications. The adapters, preselective and selective primers and AFLP amplification core mix were purchased from Applied Biosystems (Applied Biosystems, Foster City, CA, USA), while the restriction enzymes and T_4 DNA ligase were from New England Biolabs (Beverly, MA).

AFLP fragments are dominant, multilocus markers that can only be scored as present or absent, consequently any failure to amplify a fragment will reduce AFLP reliability. Preventative measures to overcome amplification failure have been included in this study, such as the use of high-quality DNA and an excess of restriction enzymes, ensuring that complete digestion occurs. Enzymes that are sensitive to DNA methylation are avoided, as restriction enzymes will not cleave methylated DNA when the restriction enzyme target site overlaps with a methylation site. High stringency PCR protocols are used guaranteeing that only primers perfectly matched to the template sequences amplify, thereby eliminating mispriming (Mueller and Wolfenbarger, 1999).

3.1.1. DNA material

A single full-sib *H. midae* F_1 family was used for DNA collection following a two-way pseudo-testcross strategy, as both parents in the cross are investigated for the source of the marker alleles. This strategy was implemented as the testcross mating configuration is unknown, unlike in a conventional testcross where the mating configuration is known beforehand and the tester is homozygous for the locus of interest. The pseudo-testcross strategy refers to a cross between heterozygous parents where one parent is heterozygous for the polymorphic markers and null in the other parent and the polymorphic markers consequently segregate in a 1:1 ratio in the progeny. The configurations of the parents are therefore only inferred after analysing the parental origin and segregation of markers in the offspring (Pérez *et al*., 2004; Liao *et al*., 2007).

DNA was extracted from 3.5-month-old abalone juveniles (refer to DNA extraction chapter, section 2.1.2.), using the extraction protocol described by Li and Guo (2004) (based on Saghai-Maroof *et al*., 1984). One hundred and twenty-six progeny were sampled at random and stored in 100% ethanol prior to DNA extraction.

3.1.2. Restriction and ligation of genomic DNA

Prior to use, the adapter pairs were annealed in a water bath at 95°C for 5 min and allowed to cool to room temperature, over a 10 min period. This annealing only needed to be carried out once (initially).

The restriction-ligation reaction (Figure 3.2) contained 5.5µl genomic DNA (~ 0.5µg/ul), 1.1µl 10x T_4 DNA ligase buffer with ATP, 1.1µl 0.5M NaCl, 0.55µl of 1mg/ml bovine serum albumin (BSA), 1U *Mse*I, 5U *Eco*RI, 1U T_4 DNA ligase, 50 pmol *Mse*I adapter, 5 pmol *Eco*RI adapter, and water to bring the final volume to 11µl. The reaction mixture was incubated overnight at room temperature to allow for complete digestion of template DNA and then diluted with 150µl $TE_{0.1}$ Buffer (20 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). The samples were stored at 2-6°C for up to 1 month, or at -15 to -25°C for longer periods.

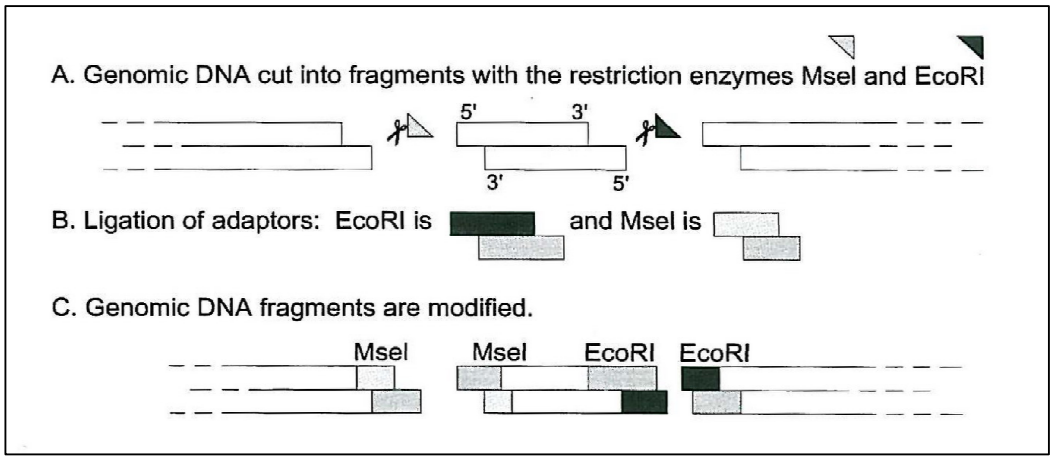


Figure 3.2. DNA template preparation (restriction reaction) and ligation of AFLP adapters (Perkin-Elmer, 1997).

3.1.3. Preselective amplification

The preselective amplification (Figure 3.3) was carried out using primers complementary to the adapter sequence and restriction enzyme recognition site with an additional nucleotide at the 3' end. The 20µl reaction contained 4µl of diluted DNA prepared by restriction-ligation, 15µl AFLP amplification core mix, and 1µl preselective primer mix (0.5µl of each *Eco*RI and *Mse*I preselective primer).

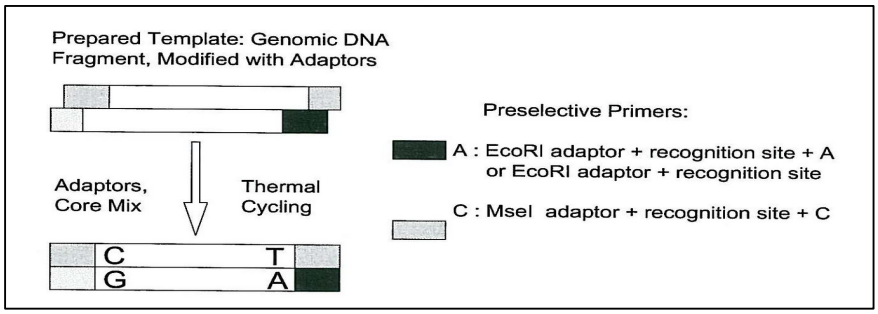


Figure 3.3. Preselective amplification of the prepared template DNA (Perkin-Elmer, 1997).

The PCR programme was: 72°C for 2 min, followed by 25 repetitive cycles of 94°C for 25 s, 56°C for 30 s, and 72°C for 2 min, with a final hold at 60°C for 30 min. All samples were stored at 2-6°C following amplification and dilution.

The amplification products' dilution factor was 10-15 fold depending on the intensity of the preselective amplification product visualised on a 1.5% (m/v) agarose gel. The amplification product (10µl product) was diluted using $TE_{0.1}$ Buffer (20 mM Tris-HCl, 0.1 mM EDTA, pH 8.0).

3.1.4. Selective PCR amplification

The selective primers are complementary to the adapter sequence (preselective primer) and contain three additional bases at the 3' end. The 5' end of the *Eco*RI selective primer has a fluorescent dye attached, fluorescently labeled with either FAM (blue), NED (yellow) or JOE (green).

The amplification reaction contained 1.5µl diluted preselective amplification reaction product, 0.5µl *Eco*RI selective dye primer containing 3 user-selected nucleotides (at 1µM), 0.5µl *Mse*I selective primer without label that contains 3 user-selected nucleotides (at 5µM), and 7.5µl AFLP amplification core mix.

The selective amplification (Figure 3.4) was run with a touchdown profile: an initial 2 min denaturing step at 94°C, followed by 10 cycles of 94°C for 20 s, 66°C for 30 s, and 72°C for 2 min, with a 1°C decrease in annealing temperature each cycle, followed by 25 cycles of amplification at 94°C for 20 s, 56°C for 30 s, and 72°C for 2 min and finally a hold at 60°C for 30 min. The samples were stored at 4°C and the selective PCR product was used for electrophoresis and data collection.

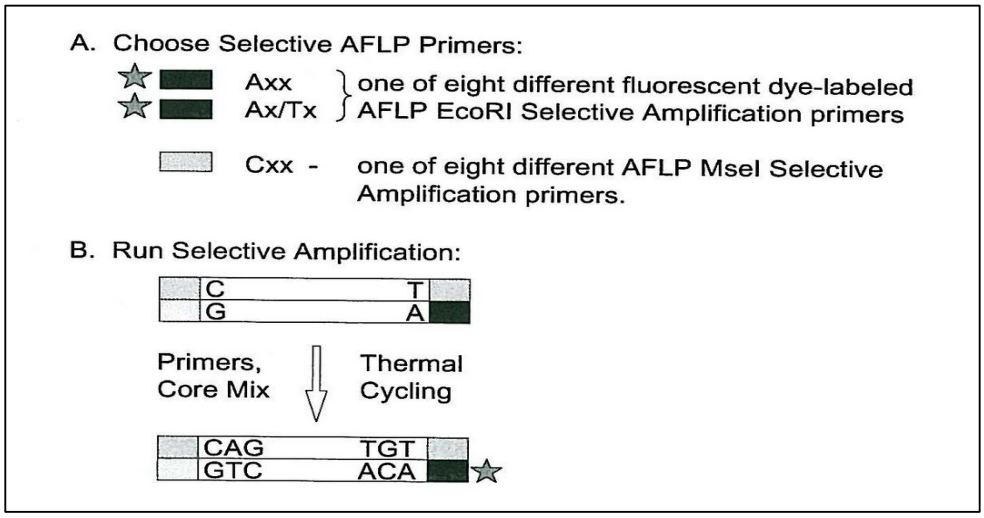


Figure 3.4. Selective amplification procedure (Perkin-Elmer, 1997).

3.1.5. Loading and marker identification of samples

Electrophoresis and data collection was carried out on an ABI 3130xl genetic analyser (Applied Biosystems, Foster City, CA, USA). The reaction mixture contained 1.0-1.5µl of selective PCR product, 12µl deionized formamide (Applied Biosystems), and 0.3µl GeneScan-500 ROX (Applied Biosystems) labeled size standard for AFLP analysis. Samples were denatured at 95^{o}C for 5 min and immediately cooled on ice before loading onto the ABI 3130xl genetic analyser in a 50cm Capillary Array (Applied Biosystems) and run in Filter Set F, following the manufacturer's instructions.

Raw data was collected using Genetic Analyser Data Collection software 3.0 (Applied Biosystems) and analysed on GeneMapper 4.0 (Applied Biosystems).

3.1.6. AFLP data collection

Fragments in the range of 50 to 500 base pairs (bp) were counted and scored. AFLP markers were scored if a peak was present in one parent and absent (null) in the other parent or present in both parents and segregating in the offspring. The peaks of the segregating AFLP loci were scored as dominant markers for mapping analysis: peak/loci present (*AA* or *Aa*) or absent (*aa*). Variations between peak intensities (dosage differences) were not used for scoring. The analysis software parameters were set according to the manufacturer's instructions (Perkin-Elmer, 1997).

The AFLP markers were named based on primer combinations used to produce them as well as their fragment size. Letters and numbers were used to code the *Eco*RI- and *Mse*I-selective

primers, respectively, followed by the letter *f* (fragment) and 2 or 3 digits representing the size in base pairs. For example, the marker *A1f267* represents the 267bp fragment produced by the primer pair, *Eco*RI primer ACA (A) and *Mse*I primer CTC (1).

3.1.7. Segregation analysis

Two kinds of segregating AFLP markers could be identified, namely type 1:1 markers, where one parent is heterozygous (peak, *A/a*) and the other is homozygous (no peak, *a/a*) and the F_1 offspring are expected to segregate in a 1:1 (peak, *A/a*: no peak, *a/a*) ratio. The other type is a 3:1 marker where both parents are heterozygous for the peak (*A/a* and *A/a*) and the offspring are expected to segregate in a 3:1 (peak, *A/a* or *A/A*: no peak, *a/a*) ratio. All segregating markers were tested for goodness of fit (observed-to-expected allelic ratios) to the 1:1 and 3:1 Mendelian ratio using chi-square analysis and a significance level of $P = 0.05$ (Appendix C). In addition, a Bonferroni correction (Rice, 1989) was applied to limit the experiment-wide error associated with multiple testing (Sokal and Rohlf, 1995). The critical χ^2 value was calculated by dividing the *P* value (0.05) by the number of linkage groups (Woram *et al.*, 2004). The use of the number of linkage groups compensates for the increased likelihood of creating type II errors when applying the Bonferroni correction to a large number of tests (Baranski *et al.*, 2006).

A marker was considered to have distorted segregation when the marker data deviated significantly from the 1:1 or 3:1 Mendelian ratios ($P < 0.05$).

3.2. Results

Among the 126 progeny sampled, 16 progeny were discarded due to poor DNA quality, and two had AFLP profiles (Figure 3.5 and 3.6) different (few scattered peaks) from those of the parents AFLP profiles (numerous peaks) (Figure 3.7). Different AFLP profiles most likely due to contaminating DNA or abalone juveniles from a different group intermingling with the sample batch. Subsequently, 108 progeny were used for AFLP marker identification and linkage analysis.

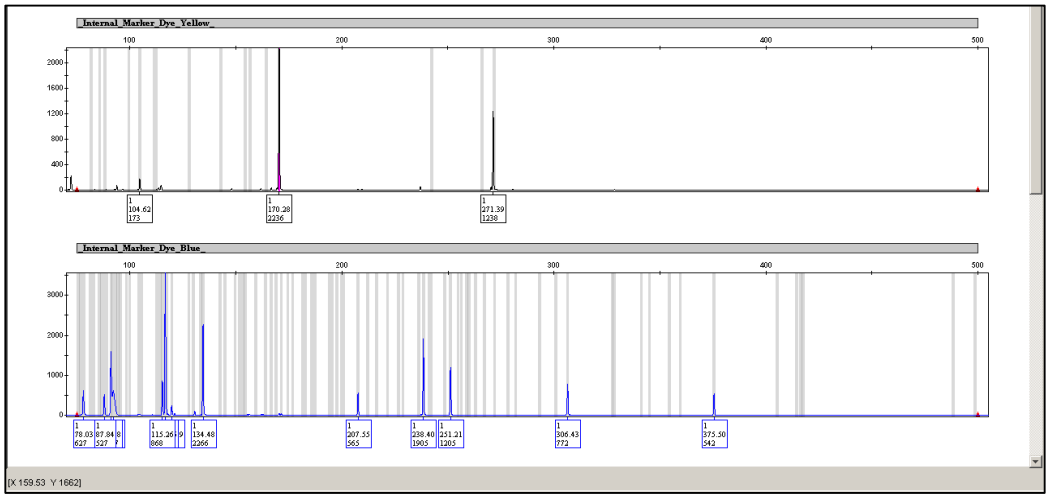


Figure 3.5. AFLP deusitometer readings of offspring sample (008) displaying different AFLP profiles compared to for example parent profiles in Figure 3.7.

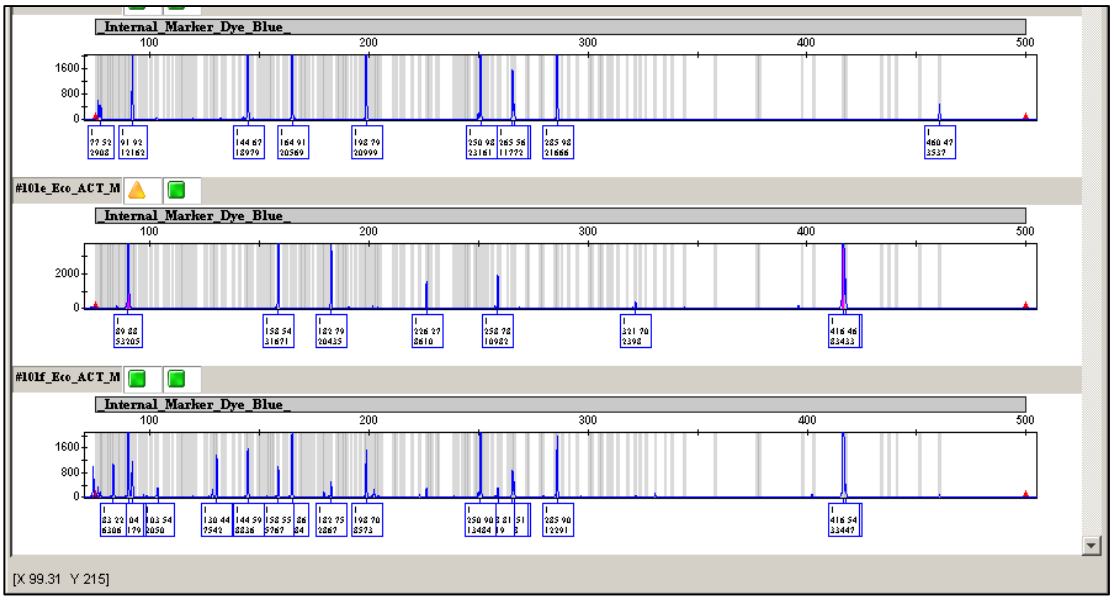


Figure 3.6. AFLP deusitometer readings of offspring sample (101) displaying different AFLP profiles compared to for example parent profiles in Figure 3.7.

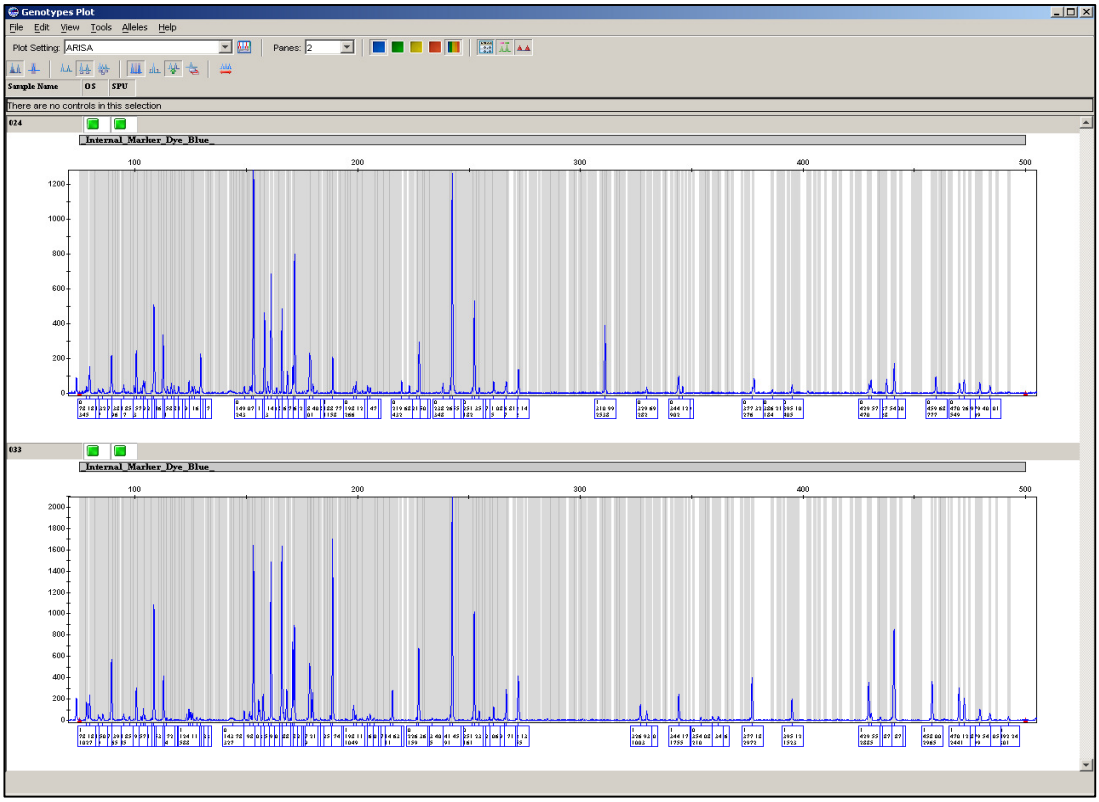


Figure 3.7. An example of an AFLP deusitometer reading of parent profiles amplified with a superior primer combination.

A total of 64 selective primer combinations were initially screened on the parents, as the progeny were not yet available at the time when primer combinations were evaluated. Primer combinations were screened for level of polymorphism and marker quality. A primer combinations' usefulness for linkage mapping was based on the total number of peaks produced, the number of polymorphic peaks present (between the parents), the amount of background noise and peak quality (intensity). A primer combination was considered superior if it had numerous polymorphic peaks, minimal background noise and good quality peaks with high peak intensity. Of the 64 primer combinations tested, 12 superior primer combinations were selected for mapping analysis (Table 3.1).

Table 3.1. AFLP primer combinations employed and polymorphic markers segregating according to a 1:1 Mendelian ratio. In parentheses (columns 1 and 2) are the letter and number codes for the *Eco*RI and *Mse*I primers, respectively.

Primer combination			**1:1 Markers**	
***Eco*RI**	***Mse*I**	**Total alleles**	**No. polymorphisms**	**% polymorphic peaks/primer**
ACA(A)	CTC(1)	66	26	39.4
ACA(A)	CTT(2)	43	19	44.2
ACT(B)	CTC(1)	40	13	32.5
ACT(B)	CAC(3)	74	35	47.3
ACT(B)	CAG(4)	56	21	37.5
ACT(B)	CAA(5)	50	25	50.0
AGC(C)	CTC(1)	36	12	33.3
ACC(D)	CTT(2)	44	24	54.5
ACC(D)	CAC(3)	36	17	47.2
ACC(D)	CTA(6)	34	7	20.6
AAG(E)	CTT(2)	47	18	38.3
AGG(F)	CTA(6)	47	24	51.1
Total		573	241	
Average		47.8	20.1	41.3

3.2.1. Markers segregating in a 1:1 Mendelian ratio

The 12 primer pairs produced a total of 573 peaks, ranging from 34 (D6) to 74 (B3) per primer pair. On average, each primer combination produced 48 peaks. Among the 573 peaks, 241 peaks were polymorphic [peak present (*A/a*) in only one parent and absent in the other parent (*a/a*) and segregating in the offspring]. Variation between the primer combinations was evident in the amount of polymorphic fragments generated. The number of polymorphic peaks present per primer pair ranged from 7 (D6) to 35 (B3) with an average of 20 peaks. The level of polymorphism among primer pairs varied, ranging from 20.6% (D6) to 54.5% (D2) (Table 3.1). Polymorphic loci segregating in a 1:1 ratio, but with exceptionally low peak intensity were not included in this study.

Among the 241 polymorphic markers showing 1:1 segregation, 173 markers segregating in either the female or male parent were selected for further analysis. The remaining 68 polymorphic markers amplified poorly in the progeny. The number of markers that segregated in the female parent was 107 (61.8%) and 66 (38.2%) markers segregated in the male parent (Table 3.2). In the female and male parent, 32 and 26 markers segregated in a 1:1 ratio (at the $P > 0.05$ level), respectively. However, following Bonferroni correction, 56 and 34 markers segregated in the female and male parent, according to 1:1 Mendelian

expectations (Appendix D), respectively. Segregation distortion remained statistically significant ($P < 0.05$) for 51 (47.7%) markers in the female and 32 (48.5%) markers in the male parent, after Bonferroni correction (Table 3.3).

Table 3.2. Analysis of markers segregating in a 1:1 Mendelian ratio.

	Total Peaks	Peaks segregating 1:1*	Peaks segregating 1:1**	Peaks segregating 1:1***	Distorted markers ($0.05 > P > 0.001$)	Distorted markers ($P < 0.001$)
Maternal parent	107	32	50	56	4	47
	100%	29.9%	46.7%	52.3%	3.7%	43.9%
Paternal parent	66	26	32	34	2	30
	100%	39.4%	48.5%	51.5%	3.0%	45.5%

* Non-significant at $P > 0.05$ level.
** Non-significant at $P > 0.01$ level.
*** Non-significant ($P > 0.05$) after Bonferroni correction.

Loci that showed statistically significant segregation distortion following Bonferroni correction at the 1:1 ratio (Table 3.3), consisted primarily of homozygote excess (null peak) of 51 markers (29.5%) compared to homozygote deficiency of 32 markers (18.5%).

Table 3.3. Markers not conforming to the 1:1 Mendelian segregation ratio.

	Total markers	Distorted markers*	Homozygous excess (+)	Homozygous deficiency (-)
Maternal parent	107	51	33	18
	100%	47.7%	30.8%	16.8%
Paternal parent	66	32	18	14
	100%	48.5%	27.3%	21.2%
Total	173	83	51	32
	100%	48.0%	29.5%	18.5%

* Displaying statistically significant segregation distortion after Bonferroni correction ($P < 0.05$).

3.2.2. Markers segregating in a 3:1 Mendelian ratio

Type 3:1 markers were observed where a peak was present in both parents and segregated in the progeny in a 3:1 ratio, which indicated that both parents were heterozygous for the marker. If the loci segregated according to Mendelian expectations ($P = 0.05$), the AFLP marker was considered polymorphic. The number of 3:1 polymorphic markers scored for mapping analysis were 164 (Table 3.4), segregating through both the female and the male parent.

Table 3.4. AFLP primer combinations and polymorphic markers segregating in a 3:1 Mendelian ratio. In parentheses (columns 1 and 2) are the letter and number codes for the *Eco*RI and *Mse*I primers, respectively.

Primer combination				3:1 Markers	
***Eco*RI**	***Mse*I**	**Total alleles**	**Peak present in both parents**	**No. polymorphisms**	**% polymorphic peaks/primer**
ACA(A)	CTC(1)	66	40	11	16.7
ACA(A)	CTT(2)	43	24	5	11.6
ACT(B)	CTC(1)	40	27	3	7.5
ACT(B)	CAC(3)	74	39	34	45.9
ACT(B)	CAG(4)	56	35	25	44.6
ACT(B)	CAA(5)	50	25	15	30.0
AGC(C)	CTC(1)	36	24	4	11.1
ACC(D)	CTT(2)	44	20	2	4.5
ACC(D)	CAC(3)	36	19	17	47.2
ACC(D)	CTA(6)	34	27	11	32.4
AAG(E)	CTT(2)	47	29	19	40.4
AGG(F)	CTA(6)	47	23	18	38.3
Total		573	332	164	
Average		47.8		13.7	27.5

On average each primer combination produced 13.7 type 3:1 markers (Table 3.4). The number of polymorphic type 3:1 markers produced per primer pair ranged from 2 (D2) to 34 (B3) with an average of 13.7. The level of polymorphism varied greatly between primer pairs, ranging from 4.5% (D2) to 47.2% (D3).

Initially, of the 332 markers analysed for 3:1 segregation, only 84 segregated according to 3:1 Mendelian expectations at $P = 0.05$. However, following Bonferroni correction, 164 markers segregated according to Mendelian expectations ($P > 0.05$) (Table 3.5; Appendix E). The 3:1 segregating markers will be used to detect any potential homologies between the two parental linkage maps.

Table 3.5. Analysis of markers segregating according to 3:1 Mendelian expectations.

	Total Peaks	**Peaks segregating 3:1***	**Peaks segregating 3:1****	**Peaks segregating 3:1*****	**Distorted markers**
Peaks segregating 3:1	332	84	131	164	168[+]
	100%	25.3%	39.5%	49.4%	50.6%

* Non-significant at $P > 0.05$ level.
** Non-significant at $P > 0.01$ level.
*** Non-significant ($P > 0.05$) following Bonferroni correction.

3.2.3. Segregation distortion

The number of polymorphic markers that showed segregation-ratio distortion ($P < 0.05$) for both 1:1 and 3:1 segregation was 83 (24.6%) (Table 3.6). Based on the 1:1 marker segregation, segregation distortion ratios in the female and male parent were 47.6% and 48.5% respectively (Table 3.2).

Table 3.6. Number of 1:1 and 3:1 segregating markers generated by AFLP amplification from the primer pairs tested and scored for linkage analysis.

	Total markers	Type 1:1 Markers	Type 3:1 markers	Distorted markers
Markers scored for linkage analysis	337	90	164	83*
	100%	26.7%	48.7%	24.6%

*Very low quality markers were not considered for mapping analysis, including markers with peaks present in both parents and not segregating in a 3:1 ratio.

3.3. Discussion

The fluorescent AFLP plant mapping kit has many advantages compared to the autoradiograph protocol originally described by Vos *et al*. (1995). The protocol is non-radioactive, the kit is predominantly PCR-based and thus less time consuming than the Vos *et al*. (1995) protocol and also more cost-effective. The fluorescent technique is automated, unlike the Vos *et al*. (1995) protocol, and highly informative. Furthermore, the automated method is reliable, reproducible and robust and can be used to produce high-density maps. The kit allows for greater accuracy when scoring data and the time saved when producing results is a great advantage. In addition, the ABI 310 genetic analyser used in this study made genotyping across samples effortless and accurate as it provided precise fragment sizing.

Disadvantages of the automated analysis are that errors may still be present, even though the optimisation of parameters increased the resolution and quality of data. For large and expanding datasets, automated analysis and scoring is the only practical option. The potential of automated analysis cannot be fully realised due to limitations of the software available. The software allows for control over a limited number of parameters, such as bin width, minimum fragment size and the amplitude threshold. The software needs to be improved to take into account the use of replicates for calibration and evaluation of the quality of the data and incorporate an automated method that will objectively and methodically choose the optimal parameters for data scoring. Despite these drawbacks, parameters that can be

controlled and set in the automated software at present, allows for repeatable, objective and much less time consuming data scoring than checking all the data manually (Meudt and Clarke, 2007). Thus, automated scoring compared to manual scoring remains the superior choice.

The reproducibility and consistency of the AFLP profiles or peaks generated using the automated AFLP protocol was tested by repeating samples at random and consistent AFLP profiles were obtained when selective amplification was successful. The reproducibility of AFLP markers has been shown in other studies (Kocher *et al*., 1998; Moore *et al*., 1999; Agresti *et al*., 2000; Naruse *et al*., 2000; Wilson *et al*., 2002; Yu and Guo, 2003; Li and Guo, 2004).

The only technical problem encountered during this study was with the *Mse*I-CTC primer, which started to produce low peak intensity AFLP profiles halfway through the study. Other studies (Yu and Guo, 2003; Li and Guo, 2004) have reported problems with poor quality of the labelled primers (*Eco*RI primer), which often resulted in low peak intensity or absence of signal. In this study three different *Eco*RI primers started exhibiting the same poor signal quality and the common denominator between these primer combinations was the *Mse*I primer, as the *Eco*RI primers used in combination with other *Mse*I primers yielded consistent results.

3.3.1. Segregation analysis

Primer combinations used in this study generated on average 20.1 polymorphic markers between the parents (type 1:1 markers). The average obtained here is slightly less than the average of 22.6 markers obtained in the Pacific oyster, *Crassostera gigas* (Li and Guo, 2004) and 23.3 markers in the eastern oyster, *C. virginica* (Yu and Guo, 2003), but is much higher than the average of 10 markers obtained in the kuruma prawn, *Penaeus japonicus* (Li *et al*., 2003). The high levels of polymorphism generally observed when using AFLP primer pairs, makes the use of these markers ideal for linkage analysis in F_1 progeny of a single cross (Li *et al*., 2005). The higher polymorphic rates observed in the other studies may be due to the fact that the animals selected for the studies were artificially selected populations (Moore *et al*., 1999; Li, Z. *et al*., 2006). In this study, the mapping family was chosen at random using a two-way pseudo-testcross strategy, as the mating configurations were not known when the family was chosen.

The female parent was more heterozygous and informative than the male parent. Of the 1:1 polymorphic markers (present in one parent, *A/a* and null in the other, *a/a*) which were used to construct the linkage maps, 107 (61.8%) segregated through the female parent, while only 66 (38.2%) segregated through the male parent (Table 3.2.). The fact that the female parent was more informative than the male parent is not an uncommon phenomenon and has been reported in other studies (Li and Guo, 2004; Liu *et al*., 2006). Longer map lengths and higher female recombination rates have been noted in numerous studies (Yu and Guo, 2003; Hubert and Hedgecock, 2004; Li and Guo, 2004; Wang *et al*., 2005; Baranski *et al*., 2006; Liu *et al*., 2006). The mechanisms resulting in the discrepancy of recombination rates and informativeness between the sexes are still not well understood, even though it is a common phenomenon. A number of explanations have been postulated for this phenomenon, such as differences between sexes in the duration of time spent in meiotic prophase, the presence of sequences which are recognised by sex-specific enzymes and differences in transcriptional activity of specific genes during meiosis between the sexes (Wang *et al*., 2004; Baranski *et al*., 2006).

Co-dominant markers can be identified by comparing differences in intensity of AFLP fragments of the same size in different individuals. The variations in intensity are thought to be positively linked with allelic copy number (Meudt and Clarke, 2007), but we were not comfortable in distinguishing co-dominant peaks in the data and did not apply this type of analysis.

3.3.2. Segregation distortion

In this study, the segregation distortion of AFLP markers is approximately 24.6%, when considering all segregating markers (1:1 and 3:1) scored for linkage analysis (Table 3.6). Departures from Mendelian expectations have been reported by other aquaculture studies using AFLP molecular markers for genetic linkage map construction: 30.53% in the guppy, *Poecilia reticulata* (Shen *et al*., 2007); 37.5% and 17.8% in Zhikong scallop, *Chlamys farreri* (Wang *et al*., 2004; Li *et al*., 2005 respectively); 26.9% in pacific oyster, *Crassostrea gigas* (Li and Guo, 2004); 16% in channel catfish, *Ictalurus punctatus* (Liu *et al*., 2003) and 13.3% in rainbow trout, *Oncorhynchus mykiss* (Young *et al*., 1998).

Segregation distortion is a problem that is frequently encountered in mapping populations used for linkage analysis (Jiang *et al*., 2000). There are several factors that may cause segregation distortion: the use of a small population for mapping; DNA that is damaged;

superior fertilisation (Tan *et al*., 2001); amplification of fragments of the same size from numerous different genomic regions (Faris *et al*., 1998; Negi *et al*., 2000); transmission that is distorted between genetically different genomes (Fishman *et al*., 2001); zygotic selection (Rick, 1969) or using parents from different populations containing a high genetic load (Fishman *et al*., 2001). Genetic load is a determination of all deleterious recessive alleles present in the gene pool of a population or in a family line. A high genetic load is therefore when a population or family line has a high frequency of harmful recessive alleles (Klug and Cummings, 2003). Segregation distortion may also be caused by a shortage in identical-by-descent homozygotes due to a high genetic load (Launey and Hedgecock, 2001) indicating selection against detrimental recessive mutations (Yu and Guo, 2003). This is due to harmful recessive alleles being filtered out of the family line, leading to a deficiency in homozygotes of the recessive alleles. Deleterious recessive alleles in homozygous form impact negatively on survival and may result in a natural selection against homozygotes of the harmful recessive alleles. However, as homozygote deficiency accounted for less of the segregation distortion (18.5%) of the 1:1 Mendelian segregation ratio identified in this study compared to homozygous excess, (29.5%) (Table 3.3) genetic load is most likely not a contributing factor to the segregation distortion observed. Variations in the level of segregation distortion may be a helpful indicator of variation within the population under investigation with regard to the populations' lineage or genetic load (Yu and Guo, 2003), but could not be measured in this study.

The segregation distortion figure identified in this study may, however, not be an accurate reflection, studies have reported selectively scoring loci, as loci with few or numerous bands were originally thought to be artefacts (Pérez *et al*., 2004). In contrast all loci (including loci with few or numerous peaks) were included in this study. Loci were not scored only if the loci contained no peaks segregating in the progeny.

3.3.3. Transferability of the AFLPs

One of the main disadvantages of AFLPs is their poor mobility between different populations and laboratories. However, the transferability between laboratories depends largely on the type of scoring systems used. Automatic genetic analysers and sequencers, such as ABI 310 used in this study, should be more transferable compared to gel-based genotyping. The best strategy for mapping in molluscs is to use AFLP markers in combination with co-dominant markers, such as microsatellites, which provide conveyable landmarks. The AFLP markers

may then saturate and fill gaps in the map. For laboratories that do not have microsatellite facilities, AFLP markers are their best choice (Li and Guo, 2004). The reasonably easy development of AFLP markers and the use of automated genotyping instead of silver staining allows for accurate sizing, which to some extent compensates for AFLP markers' poor transferability (Li *et al*., 2005).

The best strategy for developing a linkage map may be to use microsatellites in combination with AFLPs. The microsatellites can confer a backbone for the linkage map and allow some transferability of the map between laboratories and populations, and AFLPs can be used to saturate and fill the gaps of the map (Yu and Guo, 2003). This strategy should be kept in mind for the development of a dense linkage map of *H. midae* in future.

CHAPTER 4

MICROSATELLITE MARKER ANALYSIS

CHAPTER 4 - MICROSATELLITE MARKER ANALYSIS

Microsatellite (MS) markers are tandem arrays of short nucleotide repeats and are useful in the development of saturated linkage maps (Reece *et al*., 2004). Microsatellite markers compared to AFLP markers require more time and effort to develop (Baranski *et al*., 2006), but are easily transferable between populations and laboratories. Microsatellite markers are found in both coding and noncoding regions (Toth *et al*., 2000) and are predicted to be distributed throughout the genome (Cristescu *et al*., 2006). In addition, microsatellite markers have co-dominant inheritance and due to their hyper-variable nature, aid the assessments of sex-specific recombination rates in full-sib families (Hubert and Hedgecock, 2004). All these characteristics endorse microsatellite markers as model markers for linkage map development.

4.1. Materials and Methods

4.1.1. DNA material

A full-sibling *H. midae* family with 108 progeny was used for genotyping. DNA from the 3.5-month-old juveniles were extracted as described by Li and Guo (2004) (refer to DNA extraction chapter, section 2.1.2.).

4.1.2. Microsatellite markers

Microsatellite genotyping was carried out in the full-sib family using 10 microsatellite markers: HmNR106D, HmNR136D, HmNR20M, HmNR54H, HmNR120T, HmNR185D, HmNR180D, HmNR258R, HmNR281P and HmNR289P from Slabbert *et al*. (in press) and Appendix F .

The microsatellites were fluorescently labelled with either FAM (blue), VIC (green) or PET (red) (Applied Biosystems). Multiplex PCR reactions were carried out:

> **Reaction 1** was performed in a 10µl volume containing 5ng/µl genomic DNA, 1x Buffer, 2.5mM $MgCl_2$, 0.5mM dNTPs, 0.2µM of each HmNR20M and HmNR120T primer, 0.1µM of each HmNR54H primer and 0.5U *Taq* polymerase (Whitehead Scientific). The multiplex reaction was incubated at 94^{o}C for 4 min, then 35 cycles of 94^{o}C for 1 min, 52^{o}C for 10 s and 72^{o}C for 20 s, followed by a final elongation step at 72^{o}C for 10 min. The products of reaction 1 and 2 were mixed in a 1:1 ratio and genotyped simultaneously. Any remaining product was stored at 2-6^{o}C.

Reaction 2 consisted of the same PCR reagents and quantities as reaction 1 except the primers: 0.2µM of each HmNR106D primer; 0.1µM of each HmNR185D primer; and 0.05µM of each HmNR136D primer were used. The multiplex reaction was incubated at 94°C for 4 min, then 35 cycles of 94°C for 1 min, 52°C for 10 s and 72°C for 20 s, followed by a final elongation step at 72°C for 10 min. The products of reaction 1 and 2 were mixed in a 1:1 ratio and genotyped simultaneously. Any remaining product was stored at 2-6°C.

Reaction 3 was performed in a 10µl volume. The reaction contained 5ng/µl genomic DNA, 1x Buffer, 2.5mM $MgCl_2$, 0.5mM dNTPs, 0.1µM of each HmNR180D primer, 0.3µl of each HmNR258R and HmNR281P primer, 0.2µl of each HmNR289P primer and 0.5U *Taq* polymerase (Whitehead Scientific). The multiplex reaction was incubated at 94°C for 4 min, and then 35 cycles of 94°C for 1 min, 52°C for 10 s and 72°C for 20 s, followed by a final elongation step at 72°C for 15 min. The product of reaction 3 was genotyped and any remaining product was stored at 2-6°C.

4.1.3. Genotyping

The data collection and electrophoresis was conducted on an ABI 3130xl genetic analyser (Applied Biosystems). The reaction mixture contained 1.0-1.5µl of selective PCR product, 12µl deionized formamide (Applied Biosystems), and 0.3µl GeneScan-600 LIZ labeled size standard (Applied Biosystems). Samples were denatured at 95°C for 5 min and quickly cooled on ice before loading onto the ABI 3130xl genetic analyser in a 50cm Capillary Array (Applied Biosystems) and run on Filter Set G5, following the manufacturer's instructions.

Raw data was collected on Genetic Analyser Data Collection Software 3.0 (Applied Biosystems) and analysed on GeneMapper 4.0 (Applied Biosystems).

4.1.4. Segregation analysis

Segregating MS markers were evaluated for deviations from the expected 1:1 (two different alleles were detected in the parent, with one of the parents being a heterozygote) or 1:1:1:1 (four alleles were detected in the parents which were each heterozygous for a different pair of alleles) phenotypic Mendelian ratios. Segregating markers were tested with chi-square testing (χ^2) at a significance level of $P = 0.05$. Unreliable genotypes or DNA samples that failed to amplify were scored as missing data. In addition, a Bonferroni correction was applied to limit the experiment-wide error associated with multiple testing. The Bonferroni correction was

based upon the number of linkage groups [the *P* value (0.05) was divided by the number of linkage groups (*n*) under investigation, 0.05/*n*]. The number of linkage groups is deemed a more conservative and appropriate number as it compensates for the increased probability of producing type II errors when applying Bonferroni correction to large number of tests i.e. all markers segregation analysis tests (Woram *et al*., 2004).

The female and male segregating alleles were analysed independently to obtain segregation data for each sex.

A small subset (~5 samples) of the markers and samples were genotyped in duplicate to provide a measure of repeatibility.

4.2. Results

All 10 MS markers amplified successfully and were genotyped in the parents and progeny. The mapping parents were heterozygous for the same alleles in the MS marker, HmNR289P; subsequently this marker was not used for further linkage analysis (Figure 4.1).

Figure 4.1. Parents (A = female parent, B = male parent) heterozygous for the same alleles in marker HmNR289P.

Segregating null alleles were identified in two of the markers, HmNR180D and HmNR281P. The data was tested for the occurrence of null alleles making use of the methods described by Chakraborty *et al.* (1992) and Brookfield (1996). The null allele frequencies determined using the Chakraborty and Brookfield algorithms (Table 4.1) indicated that HmNR180D had a null allele present, as it had a positive value > 0.1. HmNR281P was the only other marker besides HmNR180D to generate a positive value using these methods and thus shows a stronger likelihood of containing a null allele than the other eight MS markers. These eight markers all had negative values with the Chakraborty and Brookfield 1 algorithms and zero values for the Brookfield 2 algorithm. The presence of null alleles in HmNR180D and HmNR281P were confirmed by visually analysing the genotyping data on Genemapper 4.0 (HmNR180D in Figure 4.2 and HmNR281P in Figure 4.3).

Figure 4.2. Presence of a null allele visually confirmed in microsatellite marker HmNR180D.

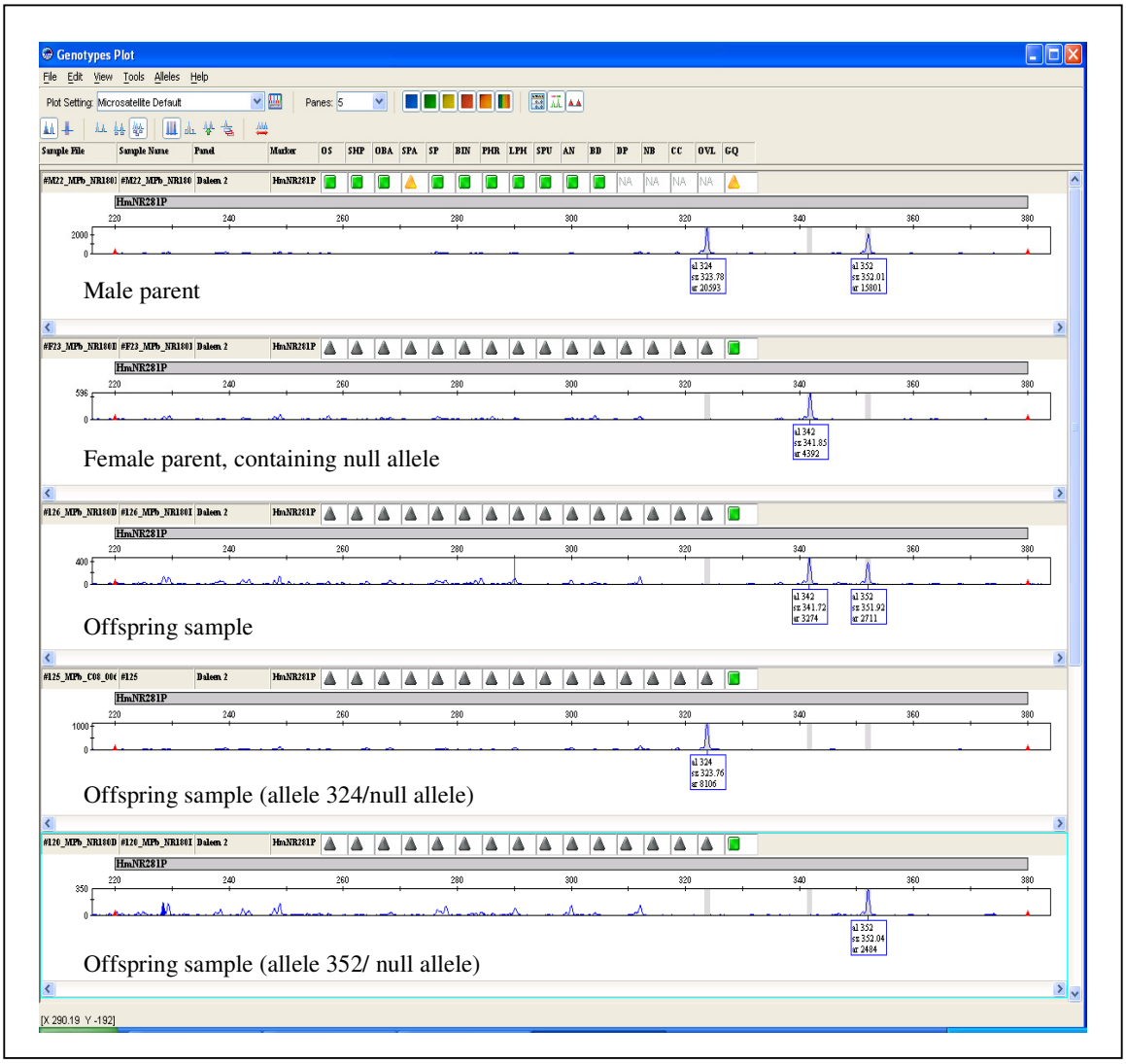


Figure 4.3. Presence of a null allele visually confirmed in microsatellite marker HmNR281P.

Table 4.1. Evaluation of estimated null allele frequencies of the MS markers using three algorithms (Chakraborty *et al*., 1992; Brookfield, 1996) in the full-sib family of this study.

Locus	Null Present	Chakraborty	Brookfield 1	Brookfield 2
HmNR106D	No	-0.1123	-0.0983	0
HmNR120T	No	-0.1467	-0.1467	0
HmNR136D	No	-0.1429	-0.0909	0
HmNR180D	Yes	0.1747	0.1167	0.1167
HmNR185D	No	-0.1436	-0.1436	0
HmNR20M	No	-0.143	-0.143	0
HmNR258R	No	-0.0683	-0.0544	0
HmNR281P	Yes	0.006	0.0047	0.0539
HmNR289P	No	-0.0655	-0.0449	0
HmNR54H	No	-0.2313	-0.2313	0

Slabbert *et al*. (in press), based on a sample group of 32 natural population individuals, calculated the null allele frequencies of the 10 MS markers used in this study using the Brookfield (1996) equation. Of the ten markers, HmNR180D and HmNR281P showed an increased likelihood of a null allele being present: HmNR180D, 0.2047 (Slabbert *et al*. in press; Table 4.2) versus 0.1167 (this study; Table 4.1); HmNR281P, 0.1066 (Slabbert *et al*. in press; Table 4.2) versus 0.0539 (this study; Table 4.1).

Table 4.2. Null allele frequencies calculated using Brookfield's (1996) equation (Slabbert *et al.* [in press]).

Locus	Null allele present	H_e	H_o	R
HmNR20M	No	0.8491	0.8519	-0.0015
HmNR54H	No	0.7574	0.7813	-0.0136
HmNR106D	No	0.8789	0.8065	0.0385
HmNR120T	No	0.9437	0.8966	0.0242
HmNR136D	No	0.7971	0.7500	0.0262
HmNR180D	Yes	0.9064	0.5161	0.2047
HmNR185D	Yes	0.8794	0.6129	0.1418
HmNR258R	No	0.7639	0.7188	0.0256
HmNR281P	Yes	0.9188	0.7143	0.1066
HmNR289P	No	0.2574	0.2500	0.0059

The nine MS markers used for linkage analysis segregated in the expected 1:1 or 1:1:1:1 Mendelian ratio, except for HmNR281P in the female parent ($P < 0.05$). However, following Bonferroni correction (based on the number of linkage groups), the deviation was not significant (Table 4.3).

Table 4.3. Segregation analysis of the nine microsatellite markers.

	Maternal	Paternal
No. of Microsatellite Markers	8	8
Conforming to either 1:1 or 1:1:1:1 segregation at the 5% level	7	8
No. of Distorted markers ($P < 0.05$)	1	0
Percentage distorted markers of total	11.1%	0%
No. of markers after Bonferroni correction	8	8

For linkage map development, separate parental maps were constructed using markers segregating according to 1:1 or 1:1:1:1 Mendelian expectations. Of the nine MS markers used for linkage analysis, eight MS markers were informative in the female parent and eight MS markers were informative in the male parent. The MS marker HmNR136D was homozygous for the female parent, consequently this marker was not informative for the maternal map and HmNR54H was homozygous for the male parent, consequently HmNR54H was not informative for the paternal map. The other seven MS (Figure 4.4 and Table 4.4) were informative in both of the parents.

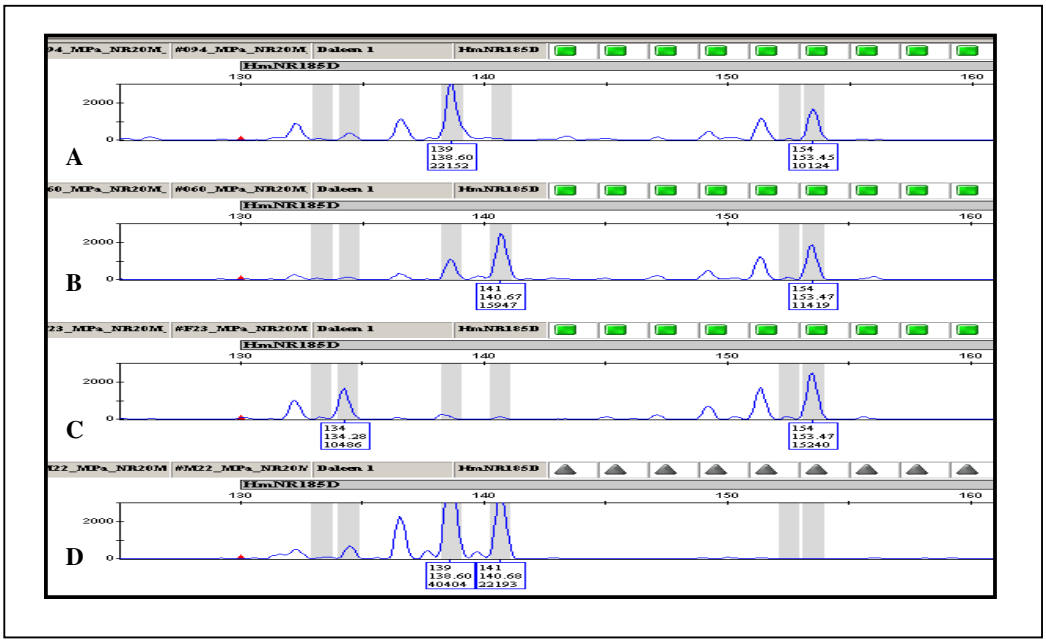


Figure 4.4. Example of polymorphic marker (HmNR185D) informative in both parents (C is female parent, D is male parent) segregating in the offspring (A and B).

Table 4.4. Microsatellite markers analysed for linkage map development.

Microsatellite markers	Accession no.	Maternal	Paternal	Ratio
HmNR20M	EF063097	ab	bc	1:1:1:1
HmNR54H	EF063103	ab	cc	1:1
HmNR106D	DQ825709	ab	bc	1:1:1:1
HmNR120T	EF121745	ab	cd	1:1:1:1
HmNR136D	DQ825710	aa	ab	1:1
HmNR180D	EF121748	ab	cd	1:1:1:1
HmNR185D	EF121750	ab	cd	1:1:1:1
HmNR258R	EF512272	ab	bc	1:1:1:1
HmNR281P	EF512274	ab	cd	1:1:1:1*
HmNR289P	EF512275	ab	ab	1:2:1[a]

[a] Heterozygous for the same alleles in both parents.
* Initial nominal deviation from Mendelian ratio ($P < 0.05$), following Bonferroni correction was non-significant ($P > 0.05$).

No genotyping errors were detected when the small subset of samples (~5 samples) was genotyped and analysed in duplicate.

4.3. Discussion

The MS marker HmNR289P that was heterozygous for the same alleles in both parents (*a/b* x *a/b*) was considered uninformative for mapping analysis and was therefore not included for linkage analysis, as the backcross model of segregation was used in the construction of the linkage map (Liao *et al*., 2007) and consequently one cannot determine from which parent allele '*a*' or '*b*' originated from in the progeny.

Without correction, approximately 11.1% of the markers segregating in the maternal parent and 0% of the markers segregating in the paternal parent appeared to segregate according to Mendelian expectations ($P < 0.05$). For the two parents the figure was 0.05%. However, following Bonferroni correction it was concluded that no significant segregation distortion occurred (Table 4.3). The level of segregation distortion of the MS markers (nine MS markers) in this study is thus very low or absent and comparable to the levels reported in other species: 0% in prawn, *Penaeus japonicus* (Moore *et al*., 1999); 11% in the eastern oyster, *Crassostrea virginica* (following Bonferroni correction) (Reece *et al*., 2004); 0% in blacklip abalone, *H. rubra* (following Bonferroni correction) (Baranski *et al*., 2006); and 5.4% in Pacific abalone, *H. discus hannai* (following Bonferroni correction) (Liu *et al*., 2006).

The cause of segregation distortion is not yet completely understood, but may be due to altered chromosome segregation or disparity in the survival and feasibility of different genotypes and gametes (Cristescu *et al*., 2006). Strong zygotic selection during the larval stages was hypothesised to cause segregation distortion in the flat oyster, *Ostrea edulis* (Naciri *et al*., 1995). This phenomenon was demonstrated experimentally by Launey and Hedgecock (2001) using *Crassostrea gigas* families, where they genotyped progeny 6 hours after fertilisation and then 2-3 months later. Their findings confirmed that segregation distortion was negligible at the zygote stage, but during development segregation distortion increased. These findings support the hypothesis that some microsatellite alleles are selected against due to their linkage to deleterious fitness gene alleles. As the progeny were sampled at 3.5-months old, it is possible that any segregation distortion present in this study (before Bonferroni correction) is a result of early zygotic viability selection. However, the number of

MS markers used in this study is very low (10 markers, although one was not used for linkage analysis) and may not be a reliable reflection of segregation distortion in this population.

Except for segregation distortion, null alleles are another shortcoming that has been noted in microsatellite loci (Gaffney *et al*., 2003; Reece *et al*., 2004). In population studies the occurrence of null alleles may confound the analysis of deviations from expected Mendelian ratios, and multiple null alleles can lead to confusion when analysing population structure. The occurrences of non-Mendelian segregation ratios of alleles have been regularly observed in numerous marine invertebrates and bivalves, which hamper the creation of a linkage map (Hare *et al*., 1996; Kocher *et al*., 1998; Brown *et al*., 2000; Launey and Hedgecock, 2001).

Despite these drawbacks, microsatellites remain ideal markers for building a foundation that will facilitate the development of higher-density linkage maps (Baranski *et al*., 2006). However, only 10 MS markers were used in this study, of which only nine were informative, and therefore this number needs to be greatly increased to ensure the anchoring of markers to linkage groups and increasing the transferability of the linkage map between laboratories and different populations.

CHAPTER 5

LINKAGE MAPPING

CHAPTER 5 – LINKAGE MAPPING

Linkage or genetic mapping has become an indispensable technique since it was first developed. Genes associated and located on the same chromosome are the basis for genetic linkage and genetic maps represent abstract models of the linear organisation of groups of genes and markers. Genetic maps are meiotic maps, as they are based on homologous recombination between loci on a chromosome. If genes or markers are located close together on a chromosome, they are generally inherited together during meiosis, as there is a link between the occurrence of recombination and physical distance between loci (Liu, 1998).

Linkage maps, particularly high-density maps, aid numerous important biological investigations: they can be used to locate genes of interest, are essential for efficient mapping of QTL and MAS, as well as comparative genome mapping (Shen *et al*., 2007).

5.1. Materials and Methods

5.1.1. Linkage analysis

Linkage analysis was carried out using the software programme MAPMAKER 3.0 (Lander and Green, 1987). The two parental maps were constructed using separate datasets of markers segregating through either the maternal or paternal parent.

The AFLP markers were scored as dominant markers and each separate parental dataset was entered into MAPMAKER 3.0 following the programme coding scheme: peak present, *Aa* (H), peak absent, *aa* (A). The F_2 backcross model was selected in preparing the data.

The microsatellite markers were scored as co-dominant markers, thus only one of the two alleles from the heterozygous parent was selected for coding and linkage analysis. For example in a cross where the parent genotypes are *AB* x *BC* (female x male), allele *A* was picked for the female parent and allele *C* for the male parent (alleles unique to each parent were chosen for coding, i.e. *A* and *C*, and not alleles shared between the parents, i.e. *B*). The F_2 backcross model was followed for the coding, similar to the AFLPs, with peak present (H) and peak absent (A). Therefore, for the female parents dataset, the allele A was scored as absent (A) and present (H) in the progeny and the same for the male parents dataset, allele C was scored as absent (A) and present (H) in the progeny. The F_2 backcross model of

MAPMAKER 3.0 assumes that all of the markers in the dataset are in coupling phase (the two alleles are located on the same sister chromatid).

All segregating markers were tested for significant deviations from the expected Mendelian segregation ratios, i.e. a 1:1 or 3:1 segregation ratio for the AFLP markers and a 1:1 or 1:1:1:1 segregation ratio for the microsatellite markers using a chi-square goodness-of-fit test (Refer chapter 3, section 3.1.7 and chapter 4, section 4.1.4). Those markers segregating at the 1:1 (marker present in the one parent, and null in the other) or 1:1:1:1 (markers where both parents are polymorphic) segregation ratios were used to construct the separate parental maps. The 3:1 segregation markers (both parents heterozygous for AFLP marker) were used to identify potential homology between the two parental maps (Table 5.1).

Table 5.1. Comparative table of the different types of segregating markers.

	1:1 segregating markers	**1:1:1:1 segregating markers**	**3:1 segregating markers**
Marker type	AFLP & Microsatellite	Microsatellite	AFLP
Parental genotypes	*Aa x aa, AB x BC*	*AB x CD*	*Aa x Aa*
F_1 genotypes	*Aa, aa Or AB, BC, BB, CA*	*AC, AD, BC, BD*	*AA, Aa, aa*
Application	Construct separate parental maps	Construct separate parental maps	Identify potential homologies between parental maps

The datasets constructed in the coupling phase were duplicated and recoded by changing H to A and A to H. This is to allow for the detection of linkage markers in repulsion phase (when the two alleles are located on different chromatids), as the linkage phase (a term which is used to designate the chromatid locations of two linked loci) is currently unknown (Cervera *et al*., 2001; Hubert and Hedgecock, 2004). Markers linked in repulsion phase had an '*r*' added at the end of their names as suggested by Lallias *et al*. (2007) to distinguish coupling and repulsion phase markers from each other; i.e. marker *B3f309r* was the recoded repulsion phase marker of the coupling phase marker *B3f309*. This coding generated two reciprocal linkage groups, and only one of these linkage groups was selected for further analysis. For example, if a group of three markers was obtained; *B3f309r*, *B3f151* and *D2f103r*, and another group of three markers *B3f309*, *B3f151r* and *D2f103*, these two groups are the

reciprocal of each other, and consequently only one of the groups was chosen at random for further linkage analysis, otherwise linkage groups with identical markers will be duplicated.

Distorted markers (deviating significantly from the 1:1 Mendelian ratio at $P = 0.05$ level; i.e. $P < 0.05$ and $P > 0.001$) were included in this study to identify possible regions of distortion. The distorted markers were suffixed with a minus (-) for homozygous deficiency and a plus (+) for homozygous excess. Only distorted markers at the $P > 0.001$ level were included in linkage mapping analysis in order to avoid possible false linkages, as distorted markers are known to influence the estimation of recombination fractions in pair-wise analysis of markers and the marker order in linkage groups (Zhu and Zhang, 2007).

Linkage groups were determined at a minimum LOD of 3.0 and a stringent maximum genetic distance between two loci of 38cM (Kosambi distance) using the GROUP command (two-point analysis) of MAPMAKER. The preliminary order of the markers in each linkage group with less than nine markers was established by the COMPARE and MAP commands (multipoint analysis). The order of markers in linkage groups with a larger number of markers ($n \geq 9$) was established using the THREE POINT or SUGGEST SUBSET, ORDER and MAP commands. Once the framework linkage groups were established with unique placement of the marker order, the relatively less stringent criteria (LOD ≥ 2.0 and ≤ 45cM genetic distance) was applied to test whether any additional markers (accessory markers) could be mapped to the framework map. Following the ordering of markers within each linkage group, the RIPPLE command was used to test the robustness of the final map order obtained. The RIPPLE function compares the likelihood of obtaining the original map order to those identified when the order of the neighbouring loci is permutated. Markers that created a conflict in map position, where there were several potential map positions with a small variation in LOD score, were placed as associated markers (Liu *et al*., 2006). All analyses were conducted with the ERROR DETECTION command on (Lincoln and Lander, 1992) selected to detect any genotyping errors. Map distance in centiMorgans was calculated using the Kosambi mapping function, which converts recombination frequencies into map distances and compensates for interference (Kosambi, 1944). The linkage groups created were numbered according to their length, in descending order, and were drawn using MapChart software (Voorrips, 2002).

Four types of markers were detected in this study: (1) framework markers (linked at LOD ≥ 3.0 and ≤ 38cM); (2) accessory markers (linked to linkage groups at LOD ≥ 2.0 and ≤ 45cM); (3) associated markers (linked but unplaced markers, as they created a conflict in map position or significantly inflated map length, where intervals became ≥50cM); and (4) unlinked markers (single unlinked markers that showed no linkage to established linkage groups).

5.1.2. Map comparison

AFLP markers (peak present in both parents, *Aa* x *Aa*) that did not deviate significantly ($P > 0.05$) from the expected 3:1 (*Aa* or *AA*:*aa*) segregation ratio (in the offspring) were used to identify any potential homologies between the two parental linkage groups. Files containing only the framework markers and the 3:1 segregating AFLP markers were loaded into MAPMAKER 3.0 software for both the maternal and paternal parent. LOD values ranging from 4.0 to 6.0 were used to add 3:1 markers individually to the framework maps using the NEAR command of MAPMAKER. Map distances were calculated from the closest framework marker and converted into Kosambi map distance (Kosambi, 1944). The linkage groups containing 3:1 segregating AFLP markers were drawn with MapChart (Voorrips, 2002) and markers homologous to both parental maps were used to identify homologies between linkage groups of the female and male parent, acting as bridging markers between the linkage groups of both maps. The segregating 3:1 markers were included in the linkage maps (shown linked to their closest framework marker i.e. map distance from closest framework marker) only if they did not contradict the map order of other 3:1 segregating markers linked to that marker and did not inflate the map length of the other linked 3:1 markers.

5.1.3. Marker distribution and genome coverage

For calculations on the marker distribution over all linkage groups, the framework, accessory and distorted markers at $0.05 > P > 0.001$ were used.

The map length and genome coverage estimations were determined on two datasets: first estimations were made using all markers, namely framework, accessory and distorted markers ($0.05 > P > 0.001$) and secondly, with only the framework markers. The map length and genome coverage estimates determined using only the framework markers will be used for further discussion, as the estimates determined using all markers may be inflated due to the

distorted and accessory markers, which were added at less stringent criteria (LOD > 2.0 and < 45cm).

5.1.3.1. Marker distribution

The correlation between the number of markers (AFLPs and microsatellites) and the length (size) of the linkage groups was analysed using the Pearson correlation coefficient (Yu and Guo, 2003; Wang *et al*., 2004; Baranski *et al*., 2006). Thereafter a *t*-test was applied to test the significance of correlation coefficient at the $P = 0.01$ level, using the null hypothesis (Ho) of no correlation.

In addition, the AFLP mapped markers were classified according to the 12 primer combinations from which they were derived (Wang *et al*., 2004). In essence, it was determined how many markers derived from each primer combination were linked in the linkage groups of both the maternal and paternal linkage maps. This consequently determined how frequently primer combinations were distributed in the male and female linkage maps in terms of markers they produced. Using these findings, it can be determined which primer combination gave rise to the most and least markers respectively in the linkage groups. Furthermore, clustering of markers generated by any primer combination in specific linkage groups or regions can be investigated.

5.1.3.2. Map length and genome coverage

The average marker spacing/intervals (s) of the two framework maps (maternal and paternal) were calculated by dividing the summed length of the map (all the linkage groups) by the number of intervals (the number of markers minus the number of linkage groups). Similarly, the average marker spacing of each linkage group was calculated by dividing the length of each linkage group by the number of intervals (the number of markers minus 1) occurring in that linkage group (Bratteler *et al*., 2006; Liu *et al*., 2006; Lallias *et al*., 2007).

Three methods were used to calculate the estimated map/genome length for each sex. First, the estimated genome length (G_{e1}) was determined by adding $2s$ (calculated above) to the length of each linkage group to account for chromosome ends (Fishman *et al*., 2001). Secondly, an estimated genome length (G_{e2}) was determined by multiplying the length of each linkage group by $(m + 1)/(m - 1)$, where m is the number of framework markers in each group (Chakravarti *et al*., 1991). These method-of-moments estimators (Hulbert *et al*., 1988; Chakravarti *et al*., 1991) have been found to be appropriate for estimating genome length

from only partial linkage data, such as was found in this study The average of the two estimates (G_{e1} and G_{e2}) was used as the estimated genome length (G_e) for *H. midae*.

The observed genome length was calculated for each parent using two estimates, one as the length of the framework map (G_{of}), and the second (G_{oa}) as the total length considering all markers (framework, accessory and distorted markers at $0.05 > P > 0.001$) (Cervera *et al*., 2001). Observed genome coverage, C_{of} (framework map coverage) and C_{oa} (total map coverage), was determined as G_{of}/G_e and G_{oa}/G_e, respectively.

5.2. Results

5.2.1. Maternal 1:1 linkage map

The linkage map of the maternal *H. midae* parent consisted of 44 framework markers (LOD ≥ 3) and 11 accessory markers (LOD ≥ 2 and ≤ 45cM) in 12 linkage groups, of which four were doublets (LOD ≥ 3) (LOD tables in Appendix G). There were 12 unlinked markers and one unplaced marker which displayed linkage to two different linkage groups in its reciprocal coding format, in linkage group 6 (*B3f273*) and in linkage group 5 (*B3f273r*) (Figure 5.1 and Table 5.2).

Table 5.2. Length, number of markers (framework and accessory), average spacing, largest and smallest intervals of linkage groups of the maternal map established with MAPMAKER 3.0. In parentheses is the number of associated markers (linked but unplaced).

Linkage group	Length (cM)	No. of framework markers	No. of accessory markers	Total no. markers	Average spacing (cM)	Largest interval (cM)	Smallest interval (cM)
1	286.0	11 (1)	3 (2)	14	28.6	47.7	14.1
2	121.4	7	1 (1)	8	20.2	38.8	2.4
3	143.4	3	2	5	35.9	43.0	24.8
4	130.0	5	1	6	26.0	40.3	12.9
5	100.7	3	1*	4	50.4	37.9	26.9
6	74.0	2	1	3	37.0	39.3	34.7
7	72.4	2	1	3	36.2	40.1	32.3
8	51.4	3	1* (1)	4	25.7	34.9	16.5
9	35.3	2	0	2	35.3	35.5	-
10	30.0	2	0	2	30.0	30.0	-
11	19.4	2	0	2	19.4	19.4	-
12	15.3	2	0	2	15.3	15.3	-
Total	1079.3	44	11	55	359.9	422.2	164.6
Average	89.9	3.7	0.8	4.5	30.0	35.2	20.6

*Linkage groups 5 and 8 contained the same marker in its reciprocal coding: *B3f273* and *B3f273r* linked at the same LOD and cM distance (not included in accessory or associated marker counts).

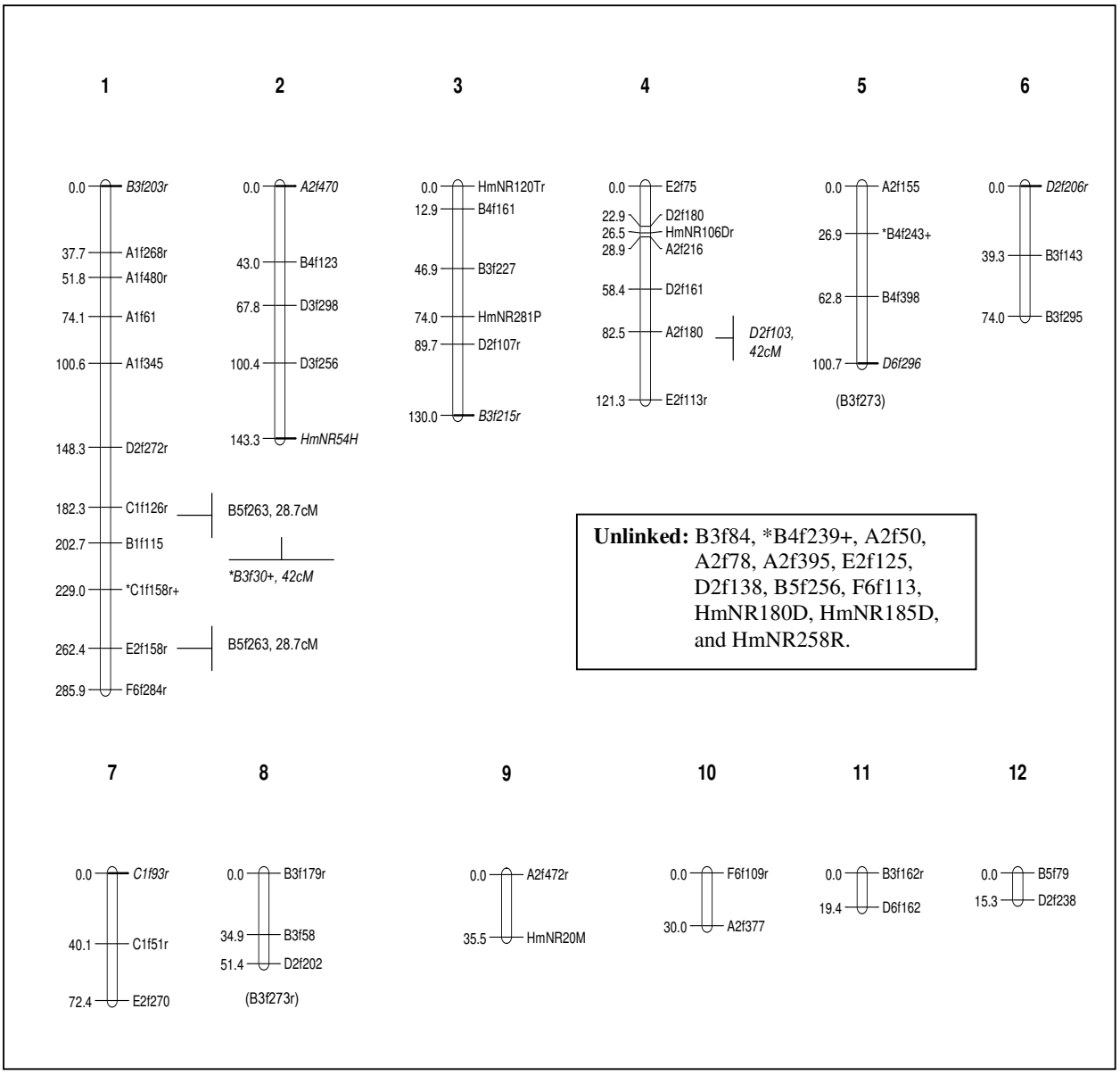


Figure 5.1. Preliminary female genetic linkage map of *H. midae* with markers indicated on the right and genetic distances (in Kosambi cM) on the left. Markers were named after their primers and fragment size, and suffixed with "-" and "+" for distorted markers (prefixed with *) showing homozygote deficiency or excess, respectively. Markers in italics are accessory markers placed on the framework map and markers shown linked to their closest framework marker are associated markers (linked but unplaced).

The linkage groups ranged in length from 15.3cM to 286cM, with an average length of 89.9cM. The average distance between the markers (framework and accessory markers) was

30cM (Table 5.2). The 12 linkage groups including the four doublets covered a total map length of 1079.3cM with the error detection command 'on', while with error detection command 'off' the total map length was 1119.3cM. This indicated that errors inflated the total map length by 40cM. The number of markers per linkage group ranged from 2 to 14, with an average of 4.5 markers per linkage group (Tables 5.2 and 5.3). The average number of framework markers per linkage group was 3.7, ranging from 2 to 11 markers per group (Table 5.2).

Table 5.3. Summary of the maternal segregating markers in *H. midae* with the distorted markers in parentheses (0.05 > *P* > 0.001, following Bonferroni correction).

Segregating markers	107 (4)
No. of markers in linkage analysis	68 (4)
Mapped AFLP markers	51 (3)
Mapped Microsatellite markers	5
Accessory markers	11
Unlinked markers	12 (1)
Linkage groups	12
Unplaced markers	5
Average no. of markers per group	4.5
Minimum no. of markers per group	2
Average marker spacing (cM)*	30
Maximum marker spacing (cM)	47.7
Minimum length of linkage group (cM)	15.3
Maximum length of linkage group (cM)	286
Total map length (cM)[a]	1079.3
Total map length (cM)[b]	1119.3

* Framework, accessory and distorted markers.
[a] Total map length with error detection on, established with MAPMAKER 3.0.
[b] Total map length with error detection off, established with MAPMAKER 3.0.

The mapped markers in the maternal map included 51 AFLP markers (including three distorted markers) and five MS markers. The 12 unlinked markers consisted of nine AFLP markers (including one distorted marker) and three MS markers (Table 5.3).

5.2.2. Paternal 1:1 linkage map

The map of the paternal *H. midae* parent consisted of 27 framework markers (LOD ≥ 3 and ≤ 38cM) and three accessory markers (LOD ≥ 2 and ≤ 45cM) in nine linkage groups, including five doublets (LOD ≥ 3). Another tentative linkage group, in the form of a doublet was

generated at LOD ≥ 2.5 and ≤ 37.2cM (Figure 5.2 and LOD tables in Appendix H). There were 14 unlinked markers (Table 5.4).

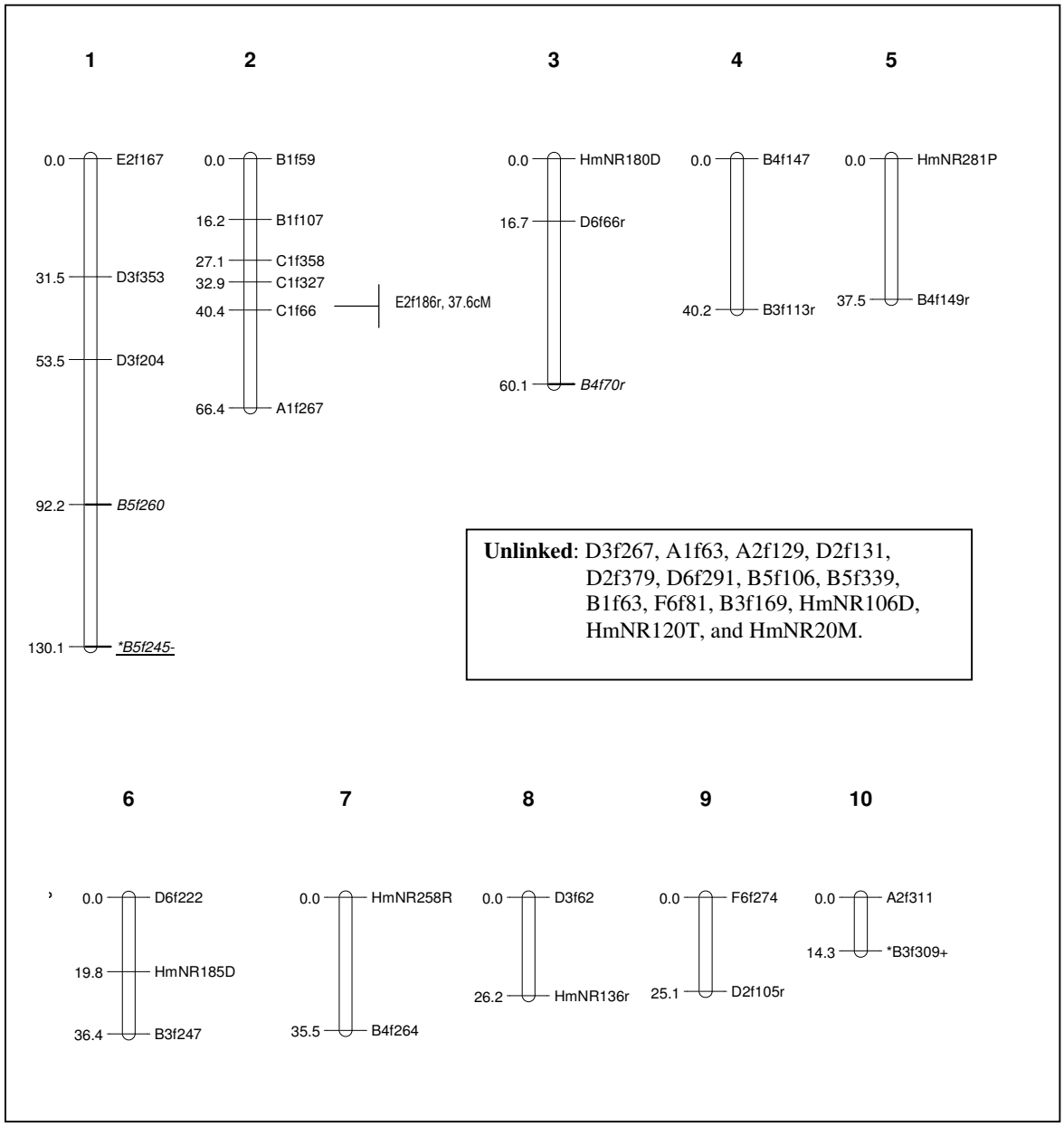


Figure 5.2. Preliminary male genetic linkage map of *H. midae* with markers indicated on the right and genetic distances (in Kosambi cM) on the left. Markers were named after their primers and fragment size, and suffixed with "-"and "+" for distorted markers (prefixed with *) showing homozygote deficiency or excess, respectively. Markers in italics are accessory markers placed on the framework map and underlined markers indicate an accessory marker linked to another accessory marker. Markers shown linked to their closest framework marker are associated markers (linked but unplaced).

Table 5.4. Length, number of markers (framework and accessory), average spacing, largest and smallest intervals of linkage groups of the paternal map established with MAPMAKER 3.0. In parentheses is the number of associated markers (linked but unplaced).

Linkage group	Length (cM)	No. of framework markers	No. of accessory markers	Total no. of markers	Average spacing (cM)	Largest interval (cM)	Smallest interval (cM)
1	130.1	3	2	5	32.5	38.7	22.0
2	66.4	7 (1)	0	7	13.3	26.0	5.8
3	60.1	2	1	3	30.1	43.5	16.7
4[a]	40.2	2	0	2	40.2	40.2	-
5	37.5	2	0	2	37.5	37.5	-
6	36.3	3	0	3	18.2	19.8	16.6
7	35.5	2	0	2	35.5	35.5	-
8	26.2	2	0	2	26.2	26.2	-
9	25.1	2	0	2	25.1	25.1	-
10	14.3	2	0	2	14.3	14.3	-
Total	471.7	27	3	30	272.8	306.8	61.1
Average	47.2	2.7	0.3	3	27.3	30.7	15.3

[a] Doublet generated at LOD ≥ 2.5 and ≤ 37.2cM.

The linkage groups ranged in length from 14.3cM to 130.1cM, with an average length of 47.2cM. The average distance between the markers (framework and accessory markers) was 27.3cM (Table 5.4). The 10 linkage groups including the six doublets covered a total map length of 471.7cM with the error detection command 'on', while with error detection command 'off' the total map length was 495cM. This indicated that errors inflated the total map length by 23.3cM. The number of markers varied from 2 to 7 per linkage group, with an average of 3 markers per linkage group (Tables 5.4 and 5.5). The average number of framework markers per linkage group was 2.7, ranging from 2 to 7 markers per group and the accessory markers per linkage group varied from 0 to 2, with an average of 0.3 (Table 5.4).

The mapped markers in the paternal map included 25 AFLP markers (including two distorted markers) and five MS markers. The 14 unlinked markers consisted of 11 AFLP markers and three MS markers and two markers showing segregation distortion ($0.05 > P > 0.001$, following Bonferroni correction) were interspersed among the markers (Table 5.5).

Table 5.5. Summary of the paternal segregating markers in *H. midae* with the distorted markers in parentheses (0.05 > *P* > 0.001, following Bonferroni correction).

Segregating markers	66 (2)
No. of markers in linkage analysis	44 (2)
Mapped AFLP markers	25 (2)
Mapped Microsatellite markers	5
Unplaced markers	1
Accessory markers	3
Unlinked markers	14
Linkage groups	10
Average no. of markers per group	3
Minimum no. of markers per group	2
Average marker spacing (cM)*	27.3
Maximum marker spacing (cM)	43.5
Minimum length of linkage group (cM)	14.3
Maximum length of linkage group (cM)	130.1
Total map length (cM)[a]	471.7
Total map length (cM)[b]	495

* Framework and accessory markers.
[a] Total map length with error detection on, established with MAPMAKER 3.0.
[b] Total map length with error detection off, established with MAPMAKER 3.0.

5.2.3. Map comparison

There was a strong indication of homology between linkage group (LG) 1 of the male map and LG 4 of the female map at a LOD ≥ 5.0 and ≤ 37.2cM. This probable homologous linkage group was based on 13 concurrent AFLP markers. The LG 1 of the male map showed tentative homology with LG 7 of the female map, but this homology was based on only four concurrent AFLP markers at a 4 ≤ LOD ≤ 5 (Figure 5.3), which is lower than the more probable homologous linkage group indicated between LG 1 (male) and LG 4 (female).

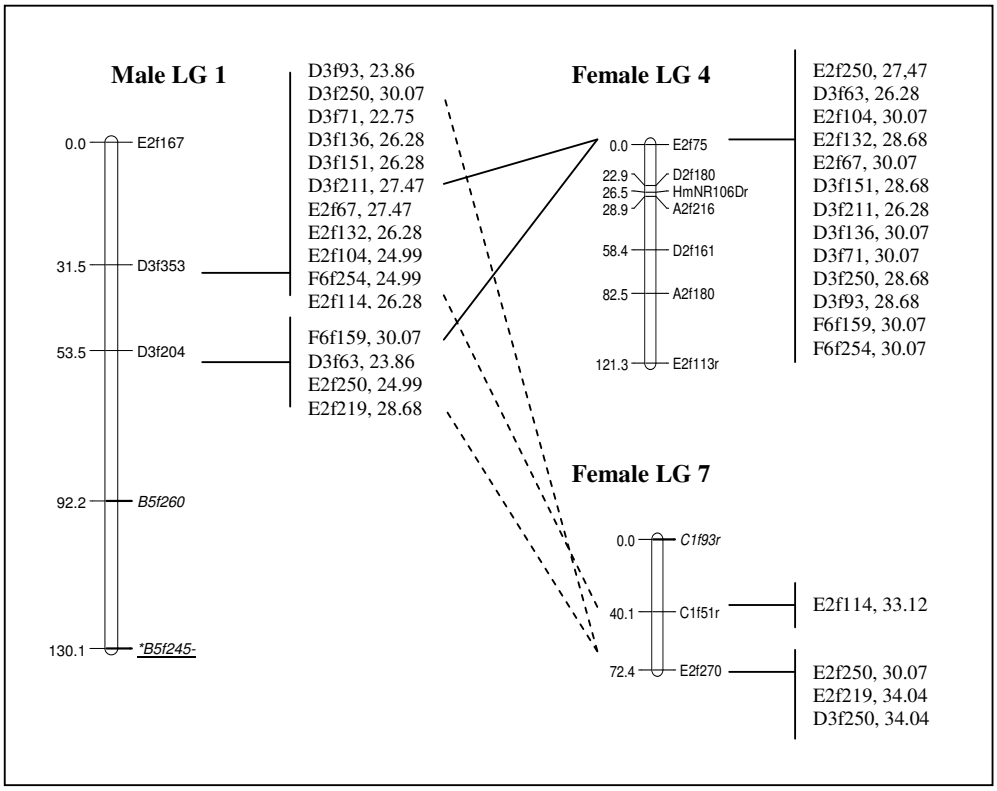


Figure 5.3. Consensus linkage groups of the male and female *H. midae* maps. Framework markers are indicated on the right; cumulative map distances (Kosambi cM) are on the left of the linkage group. Markers heterozygous in both parent, were positioned next to their closest framework marker, with map distance (in Kosambi, cM) from their closest framework marker. Lines between male and female groups indicate homologous positions.

There was further tentative homology observed between LG 3 of the female map and LG 5 of the male map, based on one concurrent microsatellite marker (HmNR281P) (Figure 5.4).

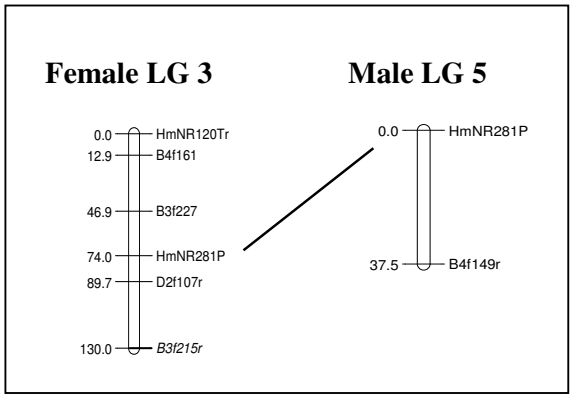


Figure 5.4. Homologous microsatellite marker, HmNR281P between LG 3 of the female map and LG 5 of the male map. Line between the female and male group indicates concurrent microsatellite position.

5.2.4. Marker distribution

The Pearson correlation coefficient analysis found a highly significant and positive correlation ($P < 0.01$) between the size (length) of the linkage group and the number of markers in the linkage group. For the male map r was calculated as 0.76 and the calculated $t = 3.31 < t_{0.01}$ rejected the null hypothesis (Ho) of no correlation. For the female map the correlation was not significant ($r = 0.95$, $t = 9.94 > t_{0.01}$). The Pearson correlation coefficient ($r = 0.95$) initially indicated that there was a highly significant positive correlation between the number of markers and linkage group size, however it was not significant according to the t-test ($P > 0.01$), indicating that there is no correlation between the number of markers and the linkage group size.

The distribution of the mapped AFLP markers classified according to the 12 primer combinations from which they were derived can be seen in Table 5.6. For the female map, the primer combinations A2, B3 and D2 appeared to be the most informative, as AFLP markers generated using these primer combinations resulted in the highest concentration of markers associated with the linkage groups. Primer combination B1 was the poorest as it had only one marker linked to the female map. Visual examination of the linkage groups of the female map in Figure 5.5 suggested that the AFLP and MS markers were distributed randomly throughout the linkage groups, however markers generated by the primer combinations A1 and D3 appeared to be clustered on linkage group 1 and linkage group 2, respectively.

Table 5.6. AFLP markers classified according to their primer combination. Primer combination coding in parentheses. [ ] are associated markers, linked but unplaced on linkage groups.

Maternal map			Paternal map		
***Eco*RI**	***Mse*I**	**No. markers**	***Eco*RI**	***Mse*I**	**No. markers**
ACA(A)	CTC(1)	4	ACA(A)	CTC(1)	1
ACA(A)	CTT(2)	6 [1]	ACA(A)	CTT(2)	1
ACT(B)	CAA(5)	1 [1]	ACT(B)	CAA(5)	2
ACT(B)	CTC(1)	1	ACT(B)	CTC(1)	2
ACT(B)	CAC(3)	8 [2]	ACT(B)	CAC(3)	3
ACT(B)	CAG(4)	4	ACT(B)	CAG(4)	4
AGC(C)	CTC(1)	4	AGC(C)	CTC(1)	3
ACC(D)	CTT(2)	7 [1]	ACC(D)	CTT(2)	1
ACC(D)	CTA(6)	2	ACC(D)	CTA(6)	2
ACC(D)	CAC(3)	2	ACC(D)	CAC(3)	3
AAG(E)	CTT(2)	4	AAG(E)	CTT(2)	1 [1]
AGG(F)	CTA(6)	2	AGG(F)	CTA(6)	1

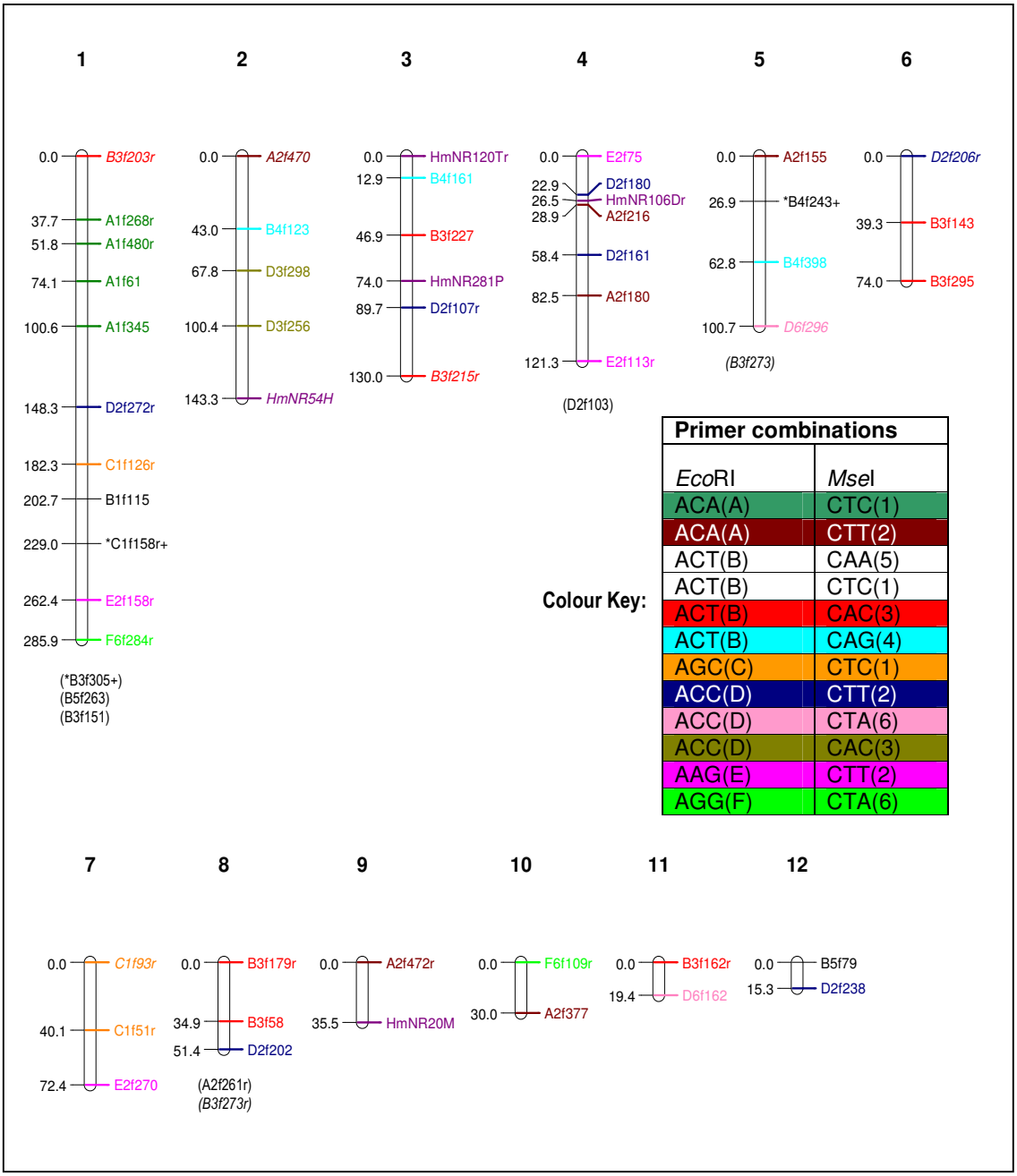


Figure 5.5. Marker distribution in the preliminary female map. Colour scheme used to indicate AFLP markers generated by the same primer combination and the microsatellite markers (HmNR-).

For the male map, the primer combinations B4, B3, C1 and D3 were the most informative, as they contributed the largest number of markers. Primer combinations A1, A2, D2 and F6 were the least informative, as only one AFLP marker of each primer combination linked to the male map. Following visual examination of the male map in Figure 5.6, primer combinations B1 and C1 clustered on linkage group 2 and B5 on linkage group 1, while the rest of the primer combinations were distributed randomly throughout the male map.

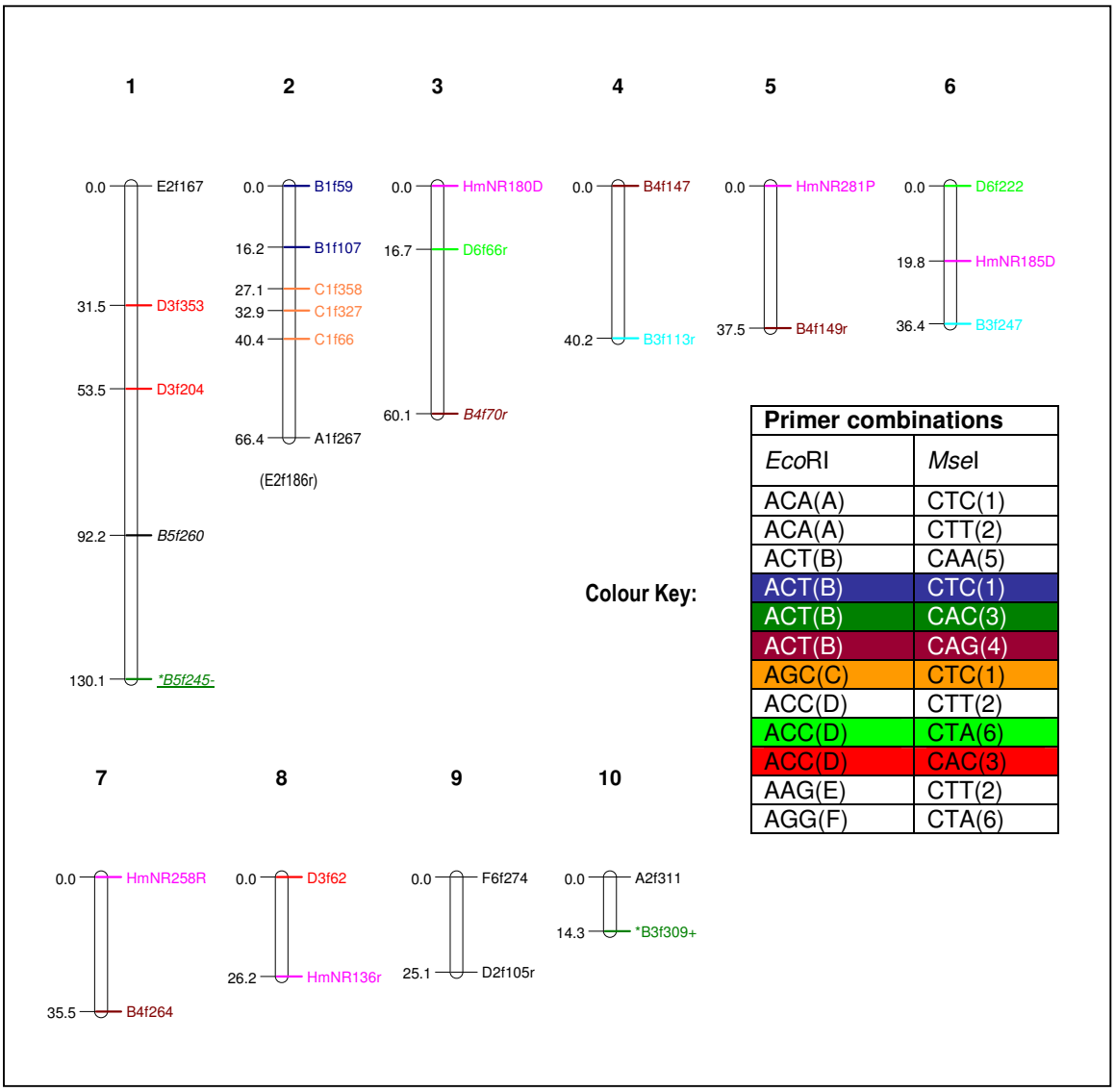


Figure 5.6. Marker distribution in the preliminary male map. Colour scheme used to indicate AFLP markers generated by the same primer combination and microsatellite markers (HmNR-).

Primer combinations were more informative in the female than male parent (Table 5.6), as a greater number of AFLP makers generated by the 12 primer combinations linked to the female map. The number of markers linked to the female map generated by the primer combinations ranged from 1 to 10 markers, while for the male map the number of markers generated by a primer combination ranged from 1 to 4.

5.2.5. Genome coverage

The genome lengths (G_e) estimated by the two methods (G_{e1} and G_{e2}) were dissimilar when determined using either all the mapped 1:1 segregating markers or only the framework markers (Table 5.7). For the G_{e1} estimate (Fishman *et al*., 2001), the average marker spacing of the maps (s) was determined both for all the markers and using only framework markers: 26.4cM for the female and 23.5cM for the male parent using all markers (framework, accessory and distorted markers) and 28.4cM for the female parent and 24.8cM for the male parent, using only the framework markers (Table 5.8).

Table 5.7. Map length and genome coverage for *H. midae*.

Map length (cM)	**Maternal**	**Paternal**	**Maternal**	**Paternal**
	All markers*		Framework markers	
Observed length				
G_{of}	819.6	312.9	819.6	312.9
G_{oa}	1079.3	471.7	1079.3	471.7
Estimated length				
G_{e1}	1760.9	968.2	1453.2	735.9
G_{e2}	2036.6	1142.5	1493.9	741.9
Average G_e	1898.8	1055.4	1473.5	738.9
Genome coverage (%)				
C_{of}	43.2	29.6	55.6	42.3
C_{oa}	56.8	44.7	73.2	63.8

* Framework, accessory and distorted markers.

Table 5.8. Average marker spacing of the maternal and paternal, *H. midae* maps.

	Average marker spacing of maps (s)
All markers*	
female	28.4
Male	24.8
Framework markers	
female	26.4
Male	23.5

*Including framework, accessory and distorted markers ($0.05 > P > 0.001$).

The estimated genome lengths, G_{e1} and G_{e2}, using only the framework markers were: 1453.2cM and 1493.9cM, respectively for the female parent and 735.9cM and 741.9cM, respectively for the male parent. The average of these two estimates (G_{e1} and G_{e2}) was used as

the expected genome length (G_e) for *H. midae*: 1473.5cM for the female and 738.9cM for the male parent (Table 5.7). Values obtained using only the framework markers were lower than the values obtained for G_{e1} and G_{e2} using all markers: 1760.9cM and 2036.6cM, respectively for the female parent and 968.2cM and 1142.5cM, respectively for the male parent. The average (G_e) of these two estimates (G_{e1} and G_{e2}) was 1898.9cM for the female and 1055.4cM for the male parent. Thus, the average genome length (G_e) was greater when determined using all markers than the G_e determined using only the framework markers.

The values for G_{of} (framework map length) and G_{oa} (total map length) were 819.6cM and 1079.3cM respectively for the female parent and 312.9cM (the length of linkage group 4 was not included in this value of G_{of}, as linkage group 4 was created at a 2.5 < LOD < 3 and does not fall under the framework marker criteria) and 471.7cM respectively for the male parent (Table 5.7 and 5.9).

Table 5.9. Length, number of markers and average spacing of linkage groups of the female and male map established with MAPMAKER 3.0 using only framework markers.

Maternal				Paternal			
Linkage group	**Length (cM)**	**No. of framework markers**	**Average marker spacing (cM)**	**Linkage group**	**Length (cM)**	**No. of framework markers**	**Average marker spacing (cM)***
1	265.4	10 (1)	29.5	**1**	54.3	3	27.2
2	121.4	7	20.2	**2**	66.4	6 (1)	13.3
3	59.0	3	29.5	**3**	17.3	2	17.3
4	90.1	5	22.5	**4***	40.2	2	40.2
5	63.5	3	31.8	**5**	37.5	2	37.5
6	33.1	2	33.1	**6**	36.3	3	18.2
7	35.5	2	35.5	**7**	35.5	2	35.5
8	51.4	3	25.7	**8**	26.2	2	26.2
9	35.5	2	35.5	**9**	25.1	2	25.1
10	30.0	2	30.0	**10**	14.3	2	14.3
11	19.4	2	19.4				
12	15.3	2	15.3				
Total	819.6	43	328	**Total**	353.1	26.0	254.7
Average	68.3	3.6	27.3	**Average**	35.3	2.6	25.5

* This linkage group was created at a 2.5 < LOD < 3 and is not included in the determination of G_{of}.

The observed genome coverage C_{of} (observed genome coverage of framework map) and C_{oa} (total map coverage, determined using all markers linked to linkage maps) varied considerably depending on which G_e estimate (estimated genome coverage) was used (Table 5.7). The values of C_{of} and C_{oa} were lower when the G_e estimate determined from all markers

was used as the observed genome coverage calculation C_{oa} takes the total map length of all the markers into account (Table 5.7). Only results of the observed genome coverage determined using the G_e of the framework markers will be discussed further in this study as the G_e of framework markers is more stringent than the G_e obtained using all the markers, because the framework markers linked to linkage maps were obtained using stricter criteria.

The observed genome coverage for C_{of} (calculated using G_{of}, which is the framework map length) was 55.6% and 42.3% for the female and male framework maps, respectively. The observed genome coverage for the respective genomes increased to 73.2% and 63.8% for the female and male parent when the C_{oa} value was determined using G_{oa}, which takes the total map length of all markers into account (Table 5.7).

5.3. Discussion

5.3.1. Linkage map

Two sex-specific linkage maps were constructed based on the male and female segregating data. As far as can be determined, this is the first time that sex-specific linkage maps have been reported for *H. midae*. The haploid genome of *H. midae* contains 18 chromosomes (Van der Merwe and Roodt-Wilding, [in press]). The female genetic map generated in this study included 12 linkage groups, while the male genetic map contained 10 linkage groups (one a doublet of LOD ≥ 2.5). The number of linkage groups obtained in this study is therefore less than the haploid chromosome number for this species. The discrepancy between the haploid chromosome number and the number of linkage groups is common for first-generation maps (Liu *et al*., 2006) and studies on other aquatic species frequently report preliminary linkage maps where the number of linkage groups are less than the haploid chromosome number, for example Kuruma prawn, *Penaeus japonicus* (Li *et al*., 2003); brown trout, *Salmo trutta* (Gharbi *et al*., 2006); and marine shrimp, *Penaeus chinensis* (Li, Z. *et al*., 2006). The low number of molecular markers used in this study (68 in the female and 44 in the male) affected the detection of recombination and coverage; subsequently the number of linkage groups obtained for the male and female parents were less than the number of chromosomes of *H. midae* (Wang *et al*., 2004; Li, Z. *et al*., 2006). To obtain a saturated linkage map of *H. midae*, considerably more molecular markers are required (Li, Z. *et al*., 2006). The population or family structure used in this study may have affected the number of linkage groups obtained (Chistiakov *et al.,* 2005), as the family was chosen at random (first full-sib family bred was

selected) following a pseudo-testcross strategy and the parents were obtained from a wild abalone population. Consequently the family chosen may not have been greatly informative for linkage mapping as their genetic origins and mating configurations are unknown.

The linkage maps presented in this study are preliminary, have relatively low molecular marker coverage and are incomplete, as indicated by the low number of linkage groups compared to the haploid chromosome number and the presence of gaps in the linkage groups (Liu *et al*., 2006; Shen *et al*., 2007). The addition of more molecular markers will increase the density and resolution of the linkage maps obtained here and the unlinked markers (12 in the female and 14 in the male parent) may display linkage patterns with the addition of more markers. Furthermore, increasing the marker density may result in the linkage groups containing only two or three markers to coalesce into larger linkage groups (Baranski *et al*., 2006).

The large intervals (> 20cM), observed in both the framework maps and the maps generated including the accessory markers, is due to the low-density of the maps obtained. The addition of markers to these maps should reduce the larger intervals and thus the gaps (Lallias *et al*., 2007).

Regions of segregation distortion were not identified in this study as so few distorted markers ($0.05 > P > 0.001$, to avoid false linkages) were included in the study: only 3 distorted markers were linked to the female map (of the 4 included for linkage analysis) and merely 2 distorted markers were linked to the male map. No clustering was observed of the distorted markers and the number of distorted markers linked to both parental linkage maps is to low to draw any conclusive results. In future, a greater number of distorted markers in both parental maps will aid the identification of the segregation distortion loci (SDL).

5.3.2. Map comparison

The 3:1 segregating markers used in the study were not distributed evenly throughout the female and male maps, which may be due to the low density of the parental maps. Potential homologies between the maternal and paternal maps (i.e. F6f159 is linked to the male map and to the female map) are given in Figure 5.3.

The probable homologous linkage groups identified in this study are not enough to establish a consensus map of *H. midae*. The addition of more framework markers and 3:1 segregating

markers to the female and male maps are required for the establishment of a consensus map. To construct a more accurate consensus map, more evenly distributed 3:1 segregating markers are needed. The addition of co-dominant markers such as microsatellites or SNPs will further increase the accuracy of the consensus map, as they will serve as anchor loci between the two parental maps. Furthermore, these markers (i.e. microsatellites and SNPs) will increase the portability of the consensus map in the context of QTL mapping (Lallias *et al*., 2007).

Consensus maps based on a combination of AFLP and microsatellite markers have been established in a number of aquaculture species, for example in tilapia, *Oreochromis niloticus* (Kocher *et al*., 1998); rainbow trout, *Oncorhynchus mykiss* (Nichols *et al*., 2003) and in common carp, *Cyprinus carpio* (Sun and Liang, 2004) and in these species have been used as starting points for the mapping of single loci and quantitative traits linked to economically important traits, as well as for studying different species (comparative analysis).

5.3.3. Marker distribution

The distribution of microsatellite and AFLP markers in the male map is relatively even (non-random distribution) as seen by the significant correlation ($P < 0.01$) between the number of markers in the linkage groups and the size (length) of the linkage groups, as determined using the Pearson correlation coefficient, while in the female map the highly positive correlation was not significant ($P > 0.01$), thus there is no correlation between the number of markers and linkage group length (size) indicating a potential random distribution between the two variables, as there is no link or association between the number of markers and the linkage group length.

A relatively even distribution is observed in the male and female maps when the linkage groups of the female and male maps are visually examined, as the microsatellites are distributed evenly throughout and only a few clusters of the AFLP markers were observed (classified according to their primer combination), of which the largest cluster consisted of only four AFLP markers (generated by the same primer combination) in the same linkage group of the female map.

In published literature, AFLP markers have been known to form clusters in some fish species (Young *et al*., 1998; Agresti *et al*., 2000; Sakamoto *et al*., 2000; Liu *et al*., 2003). However, the distribution of markers in other abalone species (on which linkage mapping analysis has been carried out to date; blacklip abalone, *H. rubra*, Baranski *et al*., 2006; Pacific abalone, *H.*

discus hannai, Liu *et al*., 2006) has been found to be relatively even. As observed in this study following visual examination of the male map and the positive and significant correlation found between the number of markers and linkage group size, indicating a parallel relationship between the two variables. In the study of Sekino and Hara (2007) some clustering in the linkage map of the Pacific abalone, *H. discus hannai* was, however, observed, as reported for the female map of this study.

Surprisingly, the reasons behind AFLP marker clustering are not yet known but several potential causes have been hypothesised, such as a large degree of variation at particular restriction sites and the presence of repeat sequences in a genomic area. It has been noted that markers tend to cluster in the region where recombination is suppressed; these regions are generally centromeres and telomeres (Shen *et al*., 2007). AFLP clustering may be a result of AFLP loci preferentially occuring in these recombinationally suppressed regions. For example, the restriction enzymes may bring about this effect, as centromeric regions commonly contain a high concentration of non-coding sequences, which contain a higher A+T content than coding sequences. The *Mse*I restriction enzyme favours the cleavage of regions with a high A+T content (Shen *et al*., 2007). The few AFLP clusters observed in the female and male map may be due to *MseI* restriction enzyme used in this study.

5.3.4. Map length and genome coverage

Two separate estimated genome lengths were determined using only the framework markers and including the accessory and distorted markers. For further discussion only those estimates determined using the framework markers (LOD $\geq$ 3 and $\leq$ 38cM) will be used, to avoid possible inflation of genome length due to the addition of the distorted and accessory markers to the map at lower LOD scores (LOD $\geq$ 2 and $\leq$ 45cM). The inflation of map length by the addition of the accessory and distorted markers can be clearly seen as the estimated genome lengths obtained using all markers (1898.9cM for the female and 1055.4cM for the male parent) is greater than those obtained using only the framework markers (1473.5cM for the female and 738.9cM for the male parent).

This study provided a first estimated genome length for *H. midae*: 1473.5cM and 738.9cM for the female and male maps respectively. The difference seen between the two maps is not an uncommon occurrence, and has been observed in vertebrates, where the male map is often shorter than the female map (Dietrich *et al*., 1996; Sakamoto *et al*., 2000). These differences have been noted in aquatic species, for example the pacific oyster, *Crassostrea gigas* (Li and

Guo, 2004); the eastern oyster, *C. virginica* (Yu and Guo, 2003) and marine shrimp, *Penaeus chinensis* (Li, Z. *et al*., 2006). Possible reasons for the variations between the estimates for each sex are: (*i*) the distribution of markers along the chromosome differs in both parents, (*ii*) recombination frequencies occurring in the female and male gametes may have differed, and (*iii*) errors within the data, such as statistical and experimental errors (Wu *et al*., 2004).

The estimated genome lengths determined in this study are similar to those obtained in the blacklip abalone, *H. rubra*, where the estimated genome lengths of the framework maps were 1586.2cM for the female and 940.5cM for the male parent (Baranski *et al*., 2006). It is, however, less than the estimated genome lengths obtained for the framework maps of the pacific abalone, *H. discus hannai*, where the genome lengths were 2584.4cM for the female and 2054.8cM for the male parent (Liu *et al*., 2006).

The genome coverage estimates (C_{of}) were 55.6% and 41.9% for the female and male framework maps respectively. The genome coverage increased to 73.2% and 56% for the female and male maps when all markers were taken into account (C_{oa}). These C_{of} estimates are somewhat less than those established in other studies, for example: 81-92% in the pacific oyster, *Crassostera gigas* (Li and Guo, 2004); ~74% in pacific abalone, *H. discus hannai* (Liu *et al*., 2006); ~75% in the blue mussel, *Mytilus edulis* (Lallias *et al*., 2007); and ~84% in the guppy, *Poecilia reticulata* (Shen *et al*., 2007). Ideally, to obtain reasonably good coverage of the genome and for the genome coverage to be of use in the mapping of genes and economically important traits i.e. in the form of QTL, > 85% genome coverage is needed. However, approximately 70% genome coverage, although low, is still potentially useful (Liu *et al*., 2006; Shen *et al*., 2007). To improve the genome coverage the addition of genetic markers is needed, especially different genetic markers, such as additional microsatellites, EST and SNPs (Wang *et al*., 2004).

CHAPTER 6

Summaries and conclusions

CHAPTER 6 – SUMMARIES AND CONCLUSIONS

6.1. DNA extraction

Genomic DNA of sufficient quality and purity for fluorescent AFLP analysis was obtained from 3.5-month-old abalone, *H. midae* juveniles. Initially, abalone larvae were considered as sources of DNA for the fluorescent AFLP analysis, as the larvae could be sampled at a early age (a few days after spawning). Following a number of tests using various extraction procedures it was apparent that the abalone larvae did not yield sufficient DNA for AFLP analysis. Subsequently, older individuals were considered, as their larger tissue mass compared to the larvae should yield larger quantities of genomic DNA. Two- and 3.5-month-old abalone juveniles were analysed and the 3.5-month-old samples yielded the highest quality DNA in sufficient quantity for fluorescent AFLP analysis. Badenhorst and Roodt-Wilding (2007) published these findings (Appendix A).

6.2. AFLP and microsatellite marker development

Fluorescent AFLP analysis was successfully carried out on an F_1 full-sib family, consisting of 108 offspring (3.5-month-old juveniles) and both parents following a pseudo-testcross mapping strategy. This was done because inbred lines of *H. midae* were not available for making testcrosses, and it would have taken a number of years before highly inbred lines are could be developed. Primer combinations were tested using only the parents, as there was a delay of 3.5-months before offspring samples could be collected. A total of 64 selective primer combinations were screened, of which 12 were selected for mapping analysis. The 12 selective primer combinations generated 241 polymorphic 1:1 segregating markers and 332 markers segregating in a 3:1 ratio (a peak present in both parents). The segregation of individual marker loci was studied using chi-square analysis and a significance level of $P = 0.05$. Ninety AFLP markers segregated according to 1:1 Mendelian expectations ($P > 0.05$ following Bonferroni correction) and 164 AFLP markers segregated according to 3:1 Mendelian expectations ($P > 0.05$ following Bonferroni correction). All that showed normal Mendelian segregation as well as those with distorted 1:1 segregation ($0.05 > P > 0.001$) were used for linkage mapping analysis.

Ten microsatellite markers identified by Slabbert *et al.* (in press) were successfully amplified in multiplex reactions and analysed. Of the 10 microsatellite markers, nine were used for

linkage mapping analysis, as one of the markers (HmNR289P) was heterozygous for the same alleles in both parents. Following segregation analysis and subsequent Bonferroni correction, it was concluded that all of the microsatellite markers showed normal Mendelian inheritance.

6.3. Linkage map development

Preliminary linkage maps were created for the female and male parents following the F_2 backcross model, as only markers heterozygous in one parent and null in the other were used. The female linkage map consisted of 12 linkage groups, while the male linkage map consisted of 10 linkage groups. The numbers of linkage groups obtained in this study are less than the haploid chromosome number of 18. This discrepancy is a common phenomenon in preliminary linkage maps (Liu *et al.*, 2006).

The female map contained 56 mapped molecular markers, while the male map had 30 mapped molecular markers. The preliminary linkage maps presented in this study are still incomplete, as can be seen from the low numbers of distinct linkage groups, the large intervals (on average > 20cM) observed in the maps and by the relatively low observed genome coverage estimates (55.6% and 41.9% for the female and male framework maps respectively). The genome coverage increased to 73.2% and 56% for the female and male maps when all markers were taken into account (C_{oa}), which is higher for the female map, but still relatively low for the male map. The estimated genome lengths of *H. midae* calculated were 1473.5cM and 738.9cM for the female and male maps, respectively.

The distribution of molecular markers was relatively even on the male map, while the female map displayed some marker clustering, and the segregating 3:1 AFLP markers were not distributed evenly. Potential homology between one of the linkage groups of the male map and two of the linkage groups of the female map was suggested by using the 3:1 segregating AFLP markers. It should be possible to merge the two parental maps, but such a merge is poorly performed using dominant markers, such as AFLPs (Li *et al.*, 2003). As only one common microsatellite marker was placed in both maps, it is therefore impossible, at this stage, to merge the two maps and obtain a consensus map.

6.4. Conclusions

The objectives and aims of this study were achieved. A preliminary linkage map was developed for *H. midae* using primarily AFLP markers. This is the first study of its kind for *H. midae* to the authors' knowledge.

Even though the linkage map developed is preliminary with relatively low genome coverage, it can serve as a basis for the development of a denser and more saturated linkage map. The addition of more molecular markers of the same type and the additional development of different markers (i.e. SNPs), as well as selecting parents and progeny which are informative for linkage mapping (i.e. in gap regions), will help facilitate the development of a complete and high-density map (Liu *et al*., 2006) of *H. midae*.

Weaknesses of this study are the low density of molecular markers mapped to both parental maps and the fact that the number of linkage groups are less than the haploid chromosome number, both attributed to the number of markers used. Furthermore, the use of juveniles means that small amounts of DNA is available for the mapping population. Whole juvenile animals were used for DNA extraction, consequently they cannot be resampled at a later stage. The strengths are that the aims of the study were achieved and it is the first study of its kind on *H. midae.* The preliminary linkage map generated in this study will serve as a foundation for the rapid development of a higher-density linkage map. In general terms, the findings of this study corroborate what has been reported in the literature, such as the common phenomenon of the female parent being more informative than the male parent, and the female map being longer than the male map (Yu and Guo, 2003; Li and Guo, 2004; Liu *et al*., 2006). The estimated genome lengths found in this study are similar to the estimated genome lengths noted by Baranski *et al*. (2006) for the blacklip abalone, *H. rubra*.

6.4.1. Future applications

In future the focus should be the expansion of the preliminary linkage map for *H. midae* based on the addition of co-dominant markers, such as microsatellite markers. The best strategy for the development of dense linkage maps is to use AFLPs in combination with microsatellites. This is due to AFLPs not being easily transferable between different laboratories and among populations as AFLPs are random markers (Coimbra *et al*., 2003) but microsatellites negate this problem as they represent transferable landmarks (Li and Guo, 2004). Ideally, microsatellites and type I markers should be used to provide a backbone for the linkage map

and AFLPs only added to fill in gaps and saturate the linkage map (Li and Guo, 2004; Yu and Guo, 2003). So as not to lose the informativeness of the AFLP markers mapped in this study, the future genome studies of *H. midae* should concentrate on identifying anchor loci located close to the AFLP markers, as this will permit the comparison of AFLP maps produced by different laboratories (Coimbra *et al*., 2003) and will compensate for the poor transferability of AFLPs.

The addition of microsatellite markers will furthermore facilitate the rapid mapping of ESTs and other type I markers and eventually lead to comparative mapping in the genus *Haliotis* and with other less related molluscs, as the microsatellite flanking regions can be used to identify and add genes to the molecular linkage map (Baranski *et al*., 2006) and type I markers characterise transcripts of genes. Comparative mapping involves the comparison of genomes of different species. The type I markers can serve as a bridge when comparing map-rich species and fairly map-poor species and facilitate the transfer of genomic information between the two (Liu and Cordes, 2004), as type I markers that are developed and added to the linkage map, can be utilised across families, and exploited in selective breeding families (Wilson *et al.,* 2002). The map will furthermore serve as a framework for gene isolation and QTL mapping (Yu and Guo, 2003). The identification of markers closely linked to desirable traits (i.e. growth, colour variation of the shell and disease resistance) is a prerequisite for the application of MAS breeding programmes, as the markers will be used as identifiers to locate the presence of desirable traits in breeding families (Wang *et al*., 2004).

Determining the position of the centromere on the linkage groups of *H. midae* will aid our understanding of the mode of recombination in the species and improve our understanding of the chromosomes of *H. midae* (i.e. structure). The location of the centromere can be identified using gene-centromere mapping techniques together with artificial triploids or gynogenetic diploids (half-tetrads). The artificial triploids and gynogenetic diploids are obtained by inhibiting the second meiotic division (Sekino and Hara, 2007). For example, in gynogenetic diploids, the centromeres of homologous chromosomes, during the first meiotic division, segregate from each other. The loci located near the centromere tend to segregate in meiosis I, while crossovers occurring between the centromere and further distal markers will result in markers segregating in meiosis II. The function of the regularity of recombination between a locus and its centromere is the proportion of second division segregation at the locus. Analysis of the four haploid products (tetrad) of a single meiotic division may indicate

how frequently a locus segregates at either the first or second meiotic division, and consequently the distance between the centromere and the gene. The gene-centromere mapping will therefore clarify the position of loci in relation to their centromere (Coimbra *et al*., 2003).

Another application includes the construction of a physical map of the *H. midae* genome, as the linkage map will serve as an anchor for the development of the physical map and provide a foundation on which contigs of overlapping clones can be gathered (Cristescu *et al*., 2006). Physical maps are commonly in the form of restriction maps of a chromosomal segment. The overlapping DNA clones are obtained by restriction enzyme digestion and the restriction fragment fingerprints are used to locate overlapping clones. If an adequate number of identical fragments are identified for any pair of clones, then the clones are likely to be overlapping. Once the fingerprints are attained for a collection of clones, a pairwise comparison for the cleavage site can be carried out. Another approach to assembling a physical map is chromosome walking. However, this is a very time consuming and laborious procedure. To identify overlapping clones in a genomic library (storage bank of cloned fragments of a genome that has been disassembled) a clone is selected and used as a probe. This step is repeated multiple times until the arrangement of a larger segment is elucidated and this assembled sequence or fragment is known as a contig (Liu, 1998). A high-density *H. midae* linkage map will provide the foundation for these contigs and necessary genomic information for the identification of the overlapping clones.

The sex-determining locus can furthermore be mapped in *H. midae* by treating the sex of the progeny as a marker. Once the sex-determining locus is identified it can be used as a probe on the chromosomes to classify sex-determining genes (Li *et al*., 2005). This was not possible in this study as the progeny were still too small to identify their sex, but in future a mapping family can be developed from which samples are only collected when the sex of the offspring can be determined. In Liu *et al*. (2006) the sex-determining locus was mapped to the male map of the Pacific abalone, *H. discus hannai*, suggesting that the sex-determining mechanism in Pacific abalone is a XY-type and the male is the heterogametic sex. No sex chromosomes have yet been observed in any abalone species. Differences in sex recombination rates are also important for future selective breeding programmes, as the sex with the lower recombination rate is expected to pass on marker-QTL associations in a much more rigid linkage fashion (Coimbra *et al.,* 2003).

In conclusion, the genetic linkage map presented here forms an ideal starting point for more detailed study of the *H. midae* genome and will provide a foundation for basic and applied studies in abalone. A high-density linkage map of *H. midae*, will facilitate the mapping of QTL of commercially important traits (i.e. growth) and for identifying the genetic foundation of evolutionary and ecologically significant traits (i.e. environmental stress and reproductive isolation). These findings will lead to the genetic enhancement of this species, which is the ultimate goal of the abalone enhancement programme initiated in 2006.

REFERENCES

REFERENCES

Agresti, J.J., Seki, S., Cnaani, A., Poompuang, S., Hallerman, E.M., Umiel, N., Hulata, G., Gall, G.A.E., May, B. 2000. Breeding new strains of tilapia: development of an artificial center of origin and linkage map based on AFLP and microsatellite loci. *Aquaculture* **185**: 43-56.

Aucamp, R. 2007. Department of Environmental Affairs and Tourism [Online]. Avalilabe: http://www.environment.gov.za/NewsMedia/MedStat/2007Apr10_1/10042007.html. [2007, November]

Badenhorst, D., Roodt-Wilding, R. 2007. Application of various DNA extraction methodologies on abalone, *Haliotis midae*, larvae and juveniles for fluorescent AFLP analysis. *Aquaculture Research* **38**: 546-549. (Appendix A)

Baranski, M., Loughnan, S., Austin, C.M., Robinson, N. 2006. A microsatellite linkage map of the blacklip abalone*, Haliotis rubra*. *Animal Genetics* **37**: 563-570.

Barkai, R., Griffiths, C.L. 1986. Diet of the South African abalone *Haliotis midae*. *South African Journal of Marine Science* **4**: 37-44.

Benton, S. 2007. South Africa bans abalone fishing. Bua News [Online]. Available: http://www.southafrica.info/ess_info/sa_glance/sustainable/abalone-261007.htm [2007, November]

Bester, A.E., Slabbert, R., D'Amato, M.E. 2004. Isolation and characterisation of microsatellite markers in the South African abalone (*Haliotis midae*). *Molecular Ecology Notes* **4**: 618-619.

Bratteler, M., Lexer, C., Widmer, A. 2006. A genetic linkage map of *Silene vulgaris* based on AFLP markers. *Genome* **49**: 320-327.

Britz, P.J. 1991. Prospects for abalone aquaculture. In Heath, R.G.M. (ed). *Aquaculture '90.* Proceedings of a joint symposium convened by the Aquaculture Association of South Africa and the University of Stellenbosch, Stellenbosch 11-13 July 1990, pp 110-113.

Brookfield, J.F.Y. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* **5**: 453-455.

Brown, B.L., Franklin, D.E., Gaffney, P.M., Hong, M., Dendanto, D., Kornfield, I. 2000. Characterisation of microsatellite loci in the eastern oyster, *Crassostrea virginica*. *Molecular Ecology* **9**: 2117-2119.

Calvo, L.M., Calvo, G.W., Burreson, E.M. 2003. Dual disease resistance in a selectively bred eastern oyster, *Crassostrea virginica*, strain tested in Chesapeake Bay. *Aquaculture* **220**: 69-87.

Campbell, D., Duchesne, P., Bernatchez L. 2003. AFLP utility for population assignment studies: analytical investigation and empirical comparison with microsatellities. *Molecular Ecology* **12**: 1979-1991.

Cartwright, D.A. 2007. Determination of marker phases in crosses with many offspring. *Genetics* **176**: 2637-2650.

Cartwright, D.A., Troggio, M., Velasco, R., Gutin, A. 2007. Genetic mapping in the presence of genotyping errors. *Genetics* **176**: 2521-2527.

Cervera, M.T., Storme, V., Ivens, B., Gusmao, J., Lliu, B.H., Hostyn, V., Slycken, J.V., Montagu, M.V., Boerjan, W. 2001. Dense genetic linkage maps of three populus species (*Populus deltoides, P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics* **158**: 787-809.

Chakraborty, R., De Andrade, M., Daiger, S.P., Budowle, B. 1992. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Annals of Human Genetics* **56**: 45-47.

Chakravarti, A., Lasher, L.K., Reefer, J.E. 1991. A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics* **128**: 175-182.

Chambers, G.K., MacAvoy, E.S. 2000. Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B* **126**: 455-476.

Chistiakov, D.A., Helleman, B., Haley, C.S., Law, A.S., Tsigenopoulos, C.S., Kotoulas, G., Bertotto, D., Libertini, A., Volckaert, F.A.M. 2005. A microsatellite linkage map of the European seabass *Dicentrarchus labrax* L. *Genetics* **170**: 1821-1826.

Coimbra, M.R.M., Kobayashi, K., Koretsugu, S., Hasegawa, O., Ohara, E., Ozaki, A., Sakamoto, T., Naruse, K., Okamoto, N. 2003. A genetic linkage map of the Japanese flounder, *Paralichthys olivaceus*. *Aquaculture* **220**: 203-218.

Cook, P. 1998. The current status of abalone farming in South Africa. *Journal of Shellfish Research* **17**: 601-602.

Cristescu, M.E.A., Colbourne, J.K., Radivojac, J., Lynch, M. 2006. A microsatellite-based genetic linkage map of the waterflea *Daphnia pulex:* On the prospect of crustacean genomics. *Genomics* **88**: 415-430.

Crollius, H.R., Jaillon, O., Dasilva, C., Ozouf-Costaz, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A., Weissenbach, J. 2000. Characterisation and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigrovirdis*. *Genome Research* **10**: 939-949.

Danzmann, R.G., Cairney, M., Davidson, W.S., Ferguson, M.M., Gharbi, K., Guyomard, R., Holm, L-E., Leder, E., Okamoto, N., Ozaki, A., Rexroad III, C.E., Sakamoto, T., Taggart, J.B., Woram, R.A. 2005. A comparative analysis of rainbow trout genome with 2 other species of fish (Artic charr and Atlantic salmon) within tetraploid derivative Salmonidae family (subfamily: Salmoninae). *Genome* **48**: 1037-1051.

Davis, G.P., Hetzel, D.J.S. 2000. Integrating molecular genetic technology with traditional approaches for genetic improvement in aquaculture species. *Aquaculture Research* **31**: 3-10.

Day, E., Branch, G.M. 2000. Evidence for a positive relationship between juvenile abalone *Haliotis midae* and the sea urchin *Parechinus angulosus* in the South-Western Cape, South Africa. *South African Journal of Marine Science* **22**: 145-156.

Dekkers, J.C. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *Journal of Animal Science* **82** (E-Suppl.): E313-E328.

Dichmont, C.M., Butterworth, D.S., Cochrane, K.I. 2000. Towards adaptive approaches to management of the South African *Haliotis midae* fishery. *South African Journal of Marine Science* **22**: 33-42.

Dietrich, W.F., Miller, J., Steen, R., Merchant, M.A., Damron-Boles, D. 1996. A comprehensive genetic map of the mouse genome. *Nature* **380**: 149-152.

Ditlecadet, D., Dufresne, F., Le François, N.R., Blier, P.U. 2006. Applying microsatellites in two commercial strains of Arctic charr (*Salvelinus alpinus*): Potential for a selective breeding program. *Aquaculture* **257**: 37-43.

Dunham, R.A., Majumdar, K., Hallerman, E., Bartley, D., Mair, G., Hulata, G., Liu, Z., Pongthana, N., Bakos, J., Penman, D., Gupta, M., Rothlisberg, P., Hoerstgen-Schwark, G. 2001. Review of the status of aquaculture genetics. In Subasinghe, R.P., Bueno, P., Phillips, M.J., Hough, C., McGladdery, S.E., Arthur, J.R. (eds). *Aquaculture in the Third Millenium.* Technical Proceedings of the Conference on Aquaculture in the Third Millennium, Bangkok, Thailand, 20-25 February 2000. NACA, Bangkok and FAO, Rome, pp 137-166.

Echt, C., Knapp, S., Liu, B.-H. 1992. Genome mapping with non-inbred crosses using Gmendel 2.0. *Maize Genetics Cooperation Newsletter* **66**: 27-29.

Elliot, N.G. 2000. Genetic improvement programmes in abalone: what is the future? *Aquaculture Research* **31**: 51-59.

Evans, B., White, R.W., Elliott, N.G. 2000. Characterization of microsatellite loci in the Australian blacklip abalone (*Haliotis rubra,* Leach). *Molecular Ecology* **9**: 1183-1184.

Falconer, D.S., Mackay, T.F.C. 1995. *Introduction to Quantitative Genetics*. (4th Ed). Addison-Wesley Longman, London.

Fallu, R. 1991. *Abalone farming*. Oxford: Fishing News Books. A division of Blackwell Scientific Publications Ltd, pp 1-181.

Fallu, R. 1994. All about Abalone. Rural Press [Online]. Available: http://www.Disease-watch.com/documents/CD/index/html/tm.htm. [2007, November]

Faris, J.D., Laddomada, B., Gill, B.S. 1998. Molecular mapping of segregation distortion loci in *Aegilops tauschii*. *Genetics* **149**: 319-327.

Ferrario, M. [Online]. Available: http://www.gastropods.com. [2007, November]

Fielding, P.J. 1995. A preliminary investigation of abalone *Haliotis midae* resources along the Transkei Coast, South Africa. *South African Journal of Marine Science* **15**: 253-261.

Fishman, L., Kelly, A.J., Morgan, E., Willis, J.H. 2001. A genetic map on the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics* **159**: 1701-1716.

Fishtech, Inc. Facts about abalone [Online]. Available: http://www.fishtech.com/html [2007, February]

Gaffney, P.M., Pierce, J.C., MacKinley, A.G., Titchen, D.A., Glenn, W.K. 2003. Pearl, a novel family of putative transposable elements in bivalve molluscs. *Journal of Molecular Evolution* **56**: 308-316.

Garcia de Leon, F.J., Canonne, M., Quillet, E., Bonhomme, F., Chatain, B. 1998. The application of microsatellite markers to breeding programmes in the sea bass, *Dicentrarchus labrax*. *Aquaculture* **159**: 303-316.

Geiger, D.L. 2000. Distribution and biogeography of the *Haliotidae* (*Gastropoda: Vetigastropoda*) world-wide. *Bollettino Malacologico* **35**: 57-120.

Gharbi, K., Gautier, A., Danzmann, R.G., Gharbi, S., Sakamoto, T., Høyheim, B., Taggart, J.B., Cairney, M., Powell, R., Krieg, F., Okamoto, N., Ferguson, M.M., Holm, L.-E., Guyomard, R. 2006. A linkage map for brown trout (*Salmo trutta*): Chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics* **172**: 2405-2419.

Gjedrem, T. 2002. Selective breeding essential for further productivity, sustainability in aquaculture. *The Advocate* 46-47.

Gjedrem, T., Fimland, E. 1995. Potential benefits from high health and genetically improved shrimp stocks. In Browdy, C.L., Hopkins, J.S. (eds). *Swimming through troubled waters*. Proceedings of the special session on shrimp farming, 1-4 February, San Diego, CA. World Aquaculture Society, Baton Rouge, LA, USA, pp 60-65.

Green, P., Falls, K., Crooks, S. 1990. CRI-MAP Documentation, version 2.4.

Gruenthal, K.M., Burton, R.S**.** 2005. Genetic diversity and species identification in the endangered white abalone (*Haliotis sorenseni*). *Conservation Genetics* **6:** 929-939.

Guo, X., Ford, S.E., De Brosse, G., Smolowitz, R., Sunila, I. 2003. Breeding and evaluation of eastern oyster strains selected for MSX, DERMO and JOD resistance (abstract). *Journal of Shellfish Research* **22**: 333-334.

Hara, M., Kikuchi, S. 1992. Increasing growth rate of abalone, *Haliotis discus hannai,* using selection techniques. *MOAA Technical Report* **106**: 21-26.

Hare, M.P., Karl, S.A., Avise, J.C. 1996. Anonymous nuclear DNA markers in the American oyster and their implications for the heterozygote deficiency phenomenon in marine bivalves. *Molecular Biology and Evolution* **13**: 334-345.

Hartl, D., Jones, E. 2001. *Genetics: Analysis of Genes and Genomes*. Jones and Bartlett Publishers, Sudbury, MA.

Hauck, M., Sweijd, N.A. 1999. A case study of abalone poaching in South Africa and its impact on fisheries management. *Journal of Marine Science* **56**: 1024-1032.

Hayes, B., He, J., Moen, T., Bennewitz, J. 2006. Use of molecular markers to maximise diversity of founder populations for aquaculture breeding programs. *Aquaculture* **255**: 573-578.

Hayes, B., Baranski, M., Goddard, M.E., Robinson, N. 2007. Optimisation of marker-assisted selection for abalone breeding programs. *Aquaculture* **265**: 61-69.

Hedgecock, D., Hubert, S., Li, G., Bucklin, K. 2002. A genetic linkage map of 100 microsatellite markers for the Pacific oyster *Crassostrea gigas* (abstract). *Journal of Shellfish Research* **21**: 381.

Huang, B., Hanna, P.J. 1998. Identification of three polymorphic microsatellite loci in blacklip abalone, *Haliotis rubra* (Leach), and detection in other abalone species. *Journal of Shellfish Research* **17**: 795-799.

Hubert, S., Hedgecock, D. 2004. Linkage maps of microsatellite DNA markers of the Pacific oyster *Crassostrea gigas*. *Genetics* **168**: 351-362.

Hulbert, S.H., Terrence, W.I., Legg, E.J., Lincoln, S.E., Lander, E.S., Michelmore, R.W. 1988. Genetic analysis of the fungus, *Bremia lactucae*, using Restriction Fragment Length Polymorphisms. *Genetics* **120**: 947-958.

Jansen, R.C., Stam, P. 1994. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**: 1447-1455.

Jiang, C., Chee, P.W., Draye, X., Morrell, P.L., Smith, C.W., Paterson, A.H. 2000. Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). *Evolution* **54**: 798-814.

Jonasson, J., Stefansson, S.E., Gudnason, A., Steinarsson, A. 1999. Genetic variation for survival and shell length of cultured red abalone (*Haliotis rufescens*) in Iceland. *Journal of Shellfish Research* **18**: 621-625.

Kawahara, I., Noro, T., Omori, M., Hasekura, O., Kijima, A. 1997. Genetic progress for growth in different selected populations of abalone *Haliotis discus hannai*, at different hatcheries. *Fish Genetics and Breeding Science* **25**: 81-90.

Ki, J.-S., Jang, G.Y., Han, M.-S. 2005. Integrated method for single-cell DNA extraction, PCR amplification, and sequencing of ribosomal DNA from harmful dinoflagellates *Cochlodinium polykrikoides* and *Alexandrium catenella*. *Marine Biotechnology* **6**: 587-593.

Kirby, V.L., Villa, R., Powers, D.A. 1998. Identification of microsatellites in the California red abalone, *Haliotis rufescens*. *Journal of Shellfish Research* **17**: 801-804.

Klug, W.S., Cummings, M.R. 2003. *Concepts of Genetics* (7th Ed). Pearson Education, Inc, pp 45-64.

Kocher, T.D., Lee, W.J., Sobolewska, H., Penman, D., McAndrew, B. 1998. A genetic linkage map of a cichlid fish, the tilapia *Oreochromis niloticus*. *Genetics* **148**: 1225-1232.

Kosambi, D.D. 1944. The estimation of map distances from recombination values. *Annals of Eugenics* **12**: 172-175.

Lallias, D., Lapègue, S., Hecquet, C., Boudry, P., Beaumont, A.R. 2007. AFLP-based genetic linkage maps of the blue mussel (*Mytilus edulis*). *Animal Genetics* **38**: 340-349.

Landau, M. 1992. *Introduction to Aquaculture*. John Wiley and Sons, Inc, pp 181-186.

Lander, E.S., Green, P. 1987. Construction of multilocus genetic linkage maps in humans. *Proceedings of the National Academy of Sciences* (*USA*) **84**: 2363-2367.

Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., Newburg, L. 1987. MAPMAKER, an interactive computer package for constructing primary genetic linkage map of experimental and natural populations. *Genomics* **1**: 174-180.

Launey, S., Hedgecock, D. 2001. High genetic load in the Pacific oyster *Crassostrea gigas*. *Genetics* **159**: 255-265.

Li, L., Guo, X. 2004. AFLP-based genetic linkage maps of the Pacific oyster *Crassostrea gigas* Thunberg. *Marine Biotechnology* **6**: 26-36.

Li, L., Xhiang, J., Liu, X., Zhang, Y., Dong, B., Zhang, X. 2005. Construction of AFLP-based genetic linkage map for Zhikong scallop, *Chlamys farreri* Jones et Preston and mapping of sex-linked markers. *Aquaculture* **245**: 63-73.

Li, Q., Yu, H., Yu, R. 2006. Genetic variability assessed by microsatellites in cultured populations of the pacific oyster (*Crassostrea gigas*) in China. *Aquaculture* **259**: 95-102.

Li, Y., Byrne, K., Miggiano, E., Whan, V., Moore, S., Keys, S., Crocos, P., Preston, N., Lehnert, S. 2003. Genetic mapping of the kuruma prawn *Penaeus japonicus* using AFLP markers. *Aquaculture* **219**: 143-156.

Li, Y., Wongprasert, K., Shekhar, M., Ryan, J., Dierens, L., Meadows, J., Preston, N., Coman, G., Lyons, R.E. 2007. Development of two microsatellite multiplex systems for black tiger shrimp *Penaeus monodon* and its application in genetic diversity study for two populations. *Aquaculture* **266**: 279-288.

Li, Z., Li, J., Wang, Q., He, Y., Liu, P. 2006. AFLP-based genetic linkage map of marine shrimp *Penaeus* (*Fenneropenaeus*) *chinensis*. *Aquaculture* **261**: 463-472.

Liao, M., Zhang, L., Yang, G., Zhu, M., Wei, Q., Zou, G., Chen, D. 2007. Development of silver carp (*Hypophthalmichthys molitrix*) and bighead carp (*Aristichthys nobilis*) genetic maps using microsatellite and AFLP markers and a pseudo-testcross strategy. *Animal Genetics* **38**: 364-370.

Lincoln, S.E., Lander, E.S. 1992. Systematic detection of errors in genetic linkage data. *Genomics* **14**: 604-610.

Lincoln, S., Daly, M., Lander, E. 1992. *Constructing Genetic Maps with MAPMAKER/EXP 3.0*. Whitehead Institute Technical Report (3rd Ed).

Liu, B-H. 1998. *Statistical Genomics: Linkage, Mapping, and QTL analysis*. Boca Raton, London, pp 9-82.

Liu, Z.J., Cordes, J.F. 2004. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* **238**: 1-37.

Liu, Z., Karsi, A., Li, P., Cao, D., Dunham, R. 2003. An AFLP-Based genetic linkage map of channel catfish (*Ictalurus punctatus*) constructed by using an interspecific hybrid resource family. *Genetics* **165**: 687-694.

Liu, X., Liu, X., Guo, X., Gao, Q., Zhao, H., Zhang, G. 2006. A preliminary genetic linkage map of the pacific abalone *Haliotis discus hannai* Ino. *Marine Biotechnology* **8**: 386-397.

Love, J.M., Knight, A.M., McAleer, M.A., Todd, J.A. 1990. Towards construction of high resolution map of the mouse genome using PCR-analyzed microsatellites. *Nucleic Acids Research* **18**: 4123-4130.

Lucas, T., Macbeth, M., Degnan, S.M., Knib, W., Degnan, B.M. 2006. Heritability estimates for growth in the tropical abalone *Haliotis asinina* using microsatellites to assign parentage. *Aquaculture* **259**: 146-152.

Lymbery, A.J., Doupe, R.G., Jenkins, G., Thorne, T. 2000. Genetics in the aquaculture industry. *Aquaculture Research* **31**: 1-2.

Macleod, F. 2007. Bye-bye abalone. Mail and Guardian [Online] Available: http://www.mg.co.za/articlePage.aspx?articleid=324475&area+/insight/insight_national/. [2007, November]

Mair, G.C. 2002. Domestication and broodstock management. *The Advocate* 39-42.

Martinez, V. 2006. *Importance and implementation of molecular markers in selective breeding programs for aquaculture species*. 8th World congress on Genetics applied to livestock production. August 13-18, 2006, Belo Horizonte, MG, Brasil.

Mayfield, S., Branch, G.M. 2000. Interrelations among rock lobsters, sea urchins, and juvenile abalone: implications for community management. *Canadian Journal of Fisheries and Aquatic Science* **57**: 2175-2185.

Meudt, H.M., Clarke, A.C. 2007. Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science* **12**: 107-117.

Mgaya, Y.D. 2000. A quantitative genetic analysis of juvenile growth for the abalone *Haliotis tuberculata* Linnaeus. In Fingerman, M., Nagabhushanam, R. (eds). *Recent advances in Marine Biotechnology*. Science Publishers, pp 59-73.

Miller, K.M., Laberee, K., Kaukinen, K.H., Li, S., Withler, R.E. 2001. Development of microsatellite loci in pinto abalone (*Haliotis kamtschatkana*). *Molecular Ecology Notes* **1**: 315-317.

Mo, C., Rinkevich, B. 2001. A simple, reliable, and fast protocol for Thraustochytrid DNA extraction. *Marine Biotechnology* **3**: 100-102.

Moore, S.S., Whan, V., Davis, G.P., Byrne, K., Hetzel, D.J.S., Preston, N. 1999. The development and application of genetic markers for the Kuruma prawn *Penaeus japonicus*. *Aquaculture* **173**: 19-32.

Mueller, U.G., Wolfenbarger, L.L. 1999. AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution* **14**: 389-394.

Naciri, Y., Vigouroux, Y., Dallas, J., Desmarais, E., Delsert, C., Bonhomme, F. 1995. Identification and inheritance of (GA/TC)n and (AC/GT)n repeats in the European flat oyster *Ostrea edulis* (L). *Molecular Marine Biology and Biotechnology* **4**: 83-89.

Naruse, K., Fukamachi, S., Mitani, H., Kondo, M., Matauoka, T., Kondo, S., Hanamura, N., Morita, Y., Hasegawa, K., Nishgaki, R., Shimda, A., Wada, H., Kusakabe, T., Suzuki, N., Kinoshita, M., Kanamori, A., Trerado, T., Kimura, H., Nonaka, M., Shima, A. 2000. A

detailed linkage map of medaka, *Oryzias latipes*, comparative genomics and genome evolution. *Genetics* **154**: 1773-1795.

Nash, C.E. 1991. Chapter 11: The production of abalone. In Nash, C.E. (ed). *Production of Aquatic animals*. Amsterdam, Oxford, New York, Tokyo, Elsevier Science Publishers B.V, pp 173-181.

Negi, M.S., Devic, M., Delseny, M., Lashmikumaran, M. 2000. Identification of AFLP fragments linked to seed coat colour in *Brassica juncea* and conversion to a SCAR marker for rapid selection. *Theoretical and Applied Genetics* **101**: 146-152.

Nichols, K.M., Young, W.P., Danzmann, R.G., Robinson, B.D., Rexroad, C., Noakes, M., Phillips, R.B., Bentzen, P., Knudsen, K., Allendorf, F.W., Cunningham, B.M., Brunelli, J., Zhang, H., Ristow, S., Drew, R., Brown, K.H., Wheeler, P.A., Thorgaard, G.H. 2003. A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). *Animal Genetics* **34**: 102-115.

Ning, Y., Liu, X., Wang, Z.Y., Guo, W., Li, Y., Xie, F. 2007. A genetic map of large yellow croaker *Pseudosciaena crocea*. *Aquaculture* **264**: 16-26.

Ohara, E., Nishimura, T., Nagakura, Y., Sakamoto, T., Mushiake, K., Okamoto, N. 2005. Genetic linkage maps of two yellowtails (*Seriola quingueradiata* and *Seriola lalandi*). *Aquaculture* **244**: 41-48.

Pérez, F., Erazo, C., Zhinaula, M., Volckaert, F., Calderón, J. 2004. A sex-specific linkage map of the white shrimp *Penaeus* (*Litopenaeus*) *vannamei* based on AFLP markers. *Aquaculture* **242**: 105-118.

Perkin-Elmer. 1995. *AFLP™ Plant Mapping Kit Protocol* (P/N 402083). Applied Biosystems.

Perkin-Elmer. 1997. *AFLP™ Plant Mapping Kit Protocol* (P/N 4303146). Applied Biosystems.

Plagányi, É.E., Butterworth, D.S. 2004. A critical look at the potential of ecopath with ecosim to assist in practical fisheries management. *African Journal of Marine Sciences* **26**: 261-287.

Qi, L., Akihiro, K. 2007. Sequences characterisation of microsatellite DNA sequences in Pacific abalone (*Haliotis discus hannai*). *Journal of Ocean University of China* **6**: 47-52.

Reece, K.S., Ribeiro, W.L., Gaffney, P.M., Carnegie, R.B., Allen, S.K. 2004. Microsatellite marker development and analysis in the eastern oyster (*Crassostrea virginica*): Confirmation of null alleles and non-Mendelian segregation ratios. *Journal of Heredity* **95**: 346-352.

Rice, W.R. 1989. Analyzing tables of statistical test. *Evolution* **43**: 223-225.

Rick, C.M. 1969. Controlled introgression of chromosomes of *Solanum pennellii* into *Lycopersicon esculentum*: segregation and recombination. *Genetics* **62**: 753-768.

Roelf, W. 2007. South Africa bids to sate Asia's abalone cravings. Reuters AlertNet [Online]. Available: http://www.alertnet.org/thenews/newsdesk.htm [2007, February]

Saghai-Maroof, M.A., Solima, K.M., Jorgenson, R.A., Allard, R.W. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences* (*USA*) **81**: 8014-8018.

Sakamoto, T., Danzmann, R.G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S.K., Woram, R.A., Okamoto, N., Ferguson, M.M., Holm, L.-E., Guyomard, R., Hoyheim, B. 2000. A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterised by large sex-specific differences in recombination rates. *Genetics* **155**: 1331-1345.

Sales, J., Britz, P.J. 2001. Research on abalone (*Haliotis midae* L.) cultivation in South Africa. *Aquaculture Research* **32**: 863-874.

Saunders, J.A., Mischke, S., Hemeida, A.A. 2001. *The use of AFLP techniques for DNA fingerprinting in plants*. Beckman Coulter, Inc, pp 1-9.

Sekino, M., Hara, M. 2001. Application of microsatellite markers to population genetics studies of Japanese Flounder *Paralichthys olivaceus*. *Marine Biotechnology* **3**: 572-589.

Sekino, M., Hara, M. 2007. Linkage maps for pacific abalone (genus *Haliotis*) based on microsatellite DNA markers. *Genetics* **175**: 945-958.

Selvamani, M.J.P., Degnan, S.M., Paetkau, D., Degnan, B.M. 2000. Highly polymorphic microsatellite loci in the Heron Reef population of the tropical abalone *Haliotis asinina*. *Molecular Ecology* **9**: 1184-1186.

Shen, X., Yang, G., Liu, Y., Liao, M., Wang, X., Zhu, M., Song, W., Zou, G., Wei, Q., Wang, D., Chen, D. 2007. Construction of genetic linkage maps of guppy (*Poecilia reticulata*) based on AFLP and microsatellite DNA markers. *Aquaculture* **271**: 178-187.

Simpson, R.J., Wilding C.S., Grahame J. 1999. Simple protocol for extracting nuclear DNA from single embryos of a marine snail. *BioTechniques* **26**: 1050-1052.

Slabbert, R., Roodt-Wilding, R. 2006. Non-destructive sampling of juvenile abalone using epipodial tentacles and mucus: method and application. *African Journal of Marine Sciences* **28**: 719-721.

Slabbert, R., Ruivo, N.R., Van den Berg, N.C., Lizamore, D.L., Roodt-Wilding, R. Isolation and characterisation of 63 microsatellite loci for the abalone, *Haliotis midae*. In press. *Journal of the World Aquaculture Society.*

Snedecor, G.W., Cochran, W.G. 1989. Chapter 10: Correlation. In Snedecor, G.W., Cochran, W.G. (ed). *Statistical Methods* (8th Ed). Iowa State University Press, Ames, Iowa, pp 177-195.

Sokal, R.R., Rohlf, J.F. 1995. *Biometry: the principles and practices of statistics in biological research* (3rd Ed). W.H. Freeman and Co., New York, pp 887.

South African Abalone [Online]. Available: http://www.web.uct.ac.za/depts/zoology/abnet/safrica.html. [2007, November]

Stam, P. 1993. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *The Plant Journal* **3**: 739-744.

Steinberg, J. 2005. The illicit abalone trade in South Africa. Occasional Paper 105 [Online]. Available: http://www.iss.co.za/pubs/papers/105/Paper105.htm [2006, October].

Sun, X., Liang, L. 2004. A genetic linkage map of common carp (*Cyprinus carpio* L.) and mapping of a locus associated with cold tolerance. *Aquaculture* **238**: 165-172.

Sweijd, N. [Online]. Available: http://www.fieldmuseum.org/…/ABALONE.jpg. [2007, November]

Sweijd, N.A., Bowie, R.C.K., Lopata, A.L., Marinaki, A.M., Harley, E.H., Cook, P.A. 1998. A PCR technique for forensic, species-level identification of abalone tissue. *Journal of Shellfish Research* **17**: 889-895.

Tan, Y.D., Wan, C., Zhu, Y., Lu, C., Xiang, Z., Deng, H.W. 2001. An amplified fragment length polymorphism map of the silkworm. *Genetics* **157**: 1277-1284.

Tang, S., Tassanakajon, A., Klinbunga, S., Jarayabhand, P., Menasveta, P. 2004. Population structure of tropical abalone (*Haliotis asinina*) in coastal waters of Thailand determined using microsatellite markers. *Marine Biotechnology* **6**: 604-611.

Tarr, R.J.Q., Williams, P.V.G., Mackenzie, A.J. 1996. Abalone, sea urchins and rock lobster: A possible ecological shift that may affect traditional fisheries. *South African Journal of Marine Science* **17**: 319-323.

Tony and Lisette Lewis Foundation. Project Highlights [Online]. Available: http://www.tllf.org.za/images/abalone.jpg. [2007, November]

Toth, G., Gaspari, Z., Jurka, J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research* **10**: 967-981.

Vadopalas, B., Bouma, J.V., Jackels, C.R., Friedman, C.S. 2006. Application of real-time PCR for simultaneous identification and quantification of larval abalone. *Journal of Experimental Marine Biology and Ecology* **334**: 219-228.

Van der Merwe, M., Roodt-Wilding, R. Chromosome number of South Africa abalone *Haliotis midae*. In press. *African Journal of Marine Science.*

Van Eenennaam, A. 2004. "Marker-Assisted Selection Backgrounder". Sierra Foothill Research and Extension Center. Paper 2004_marker-assisted_selection_backgrounder. Available: http://repositories.cdlib.org/anrrec/sfrec/2004_marker-assisted_selection_backgrounder. [2007, July]

Voorrips, R.E. 2002. MapChart: software for the graphical presentation of linkage maps and QTL. *The Journal of Heredity* **93**: 77-78.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van der Lee, T., Hornes, A.F., Pot, J., Peleman, J., Kuiper, M., Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**: 4407-4414.

Walsh, P.S., Metzger, D.A., Higuchi, R. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* **10**: 506-513.

Wang, L., Song, L., Chang, Y., Xu, W., Ni, D., Guo, X. 2005. A preliminary genetic map of Zhikong scallop (*Chlamys farreri). Aquaculture Research* **36**: 643-653.

Wang, S., Bao, W., Pan, J., Zhang, L., Yao, B., Zhan, A., Bi, K., Zhang, Q. 2004. AFLP linkage map of an intraspecific cross in *Chlamys farreri*. *Journal of Shellfish Research* **23**: 491-499.

Ward, R.D., English, L.J., McGoldrick, D.J., Maquire, G.B., Nell, J.A., Thompson, P.A. 2000. Genetic improvement of the pacific oyster *Crassostrea gigas* (Thunberg) in Australia. *Aquaculture Research* **31**: 35-44.

Weisstein, E.W. "Bonferroni Correction." From MathWorld – A Wolfram Web Resource [Online]. Available: http://mathworld.wolfram.com/BonferroniCorrection.html. [2007, November]

Wilson, K., Li, Y., Whan, V., Lehert, S., Byrne, K., Moore, S., Pongsomboon, S., Tassanakajon, A., Rosenberg, G., Ballment, E., Fayazi, Z., Swan, J., Kenway, M., Benzie, J. 2002. Genetic mapping of the black tiger shrimp *Penaeus monodon* with amplified fragment length polymorphism. *Aquaculture* **204**: 297-309.

Wood, A.D., Buxton, C.D. 1996a. Aspects of the biology of the abalone *Haliotis midae* (Linne, 1758) on the East coast of South Africa.1. Feeding biology. *South African Journal of Marine Science* **17**: 61-68.

Wood, A.D., Buxton, C.D. 1996b. Aspects of the biology of the abalone *Haliotis midae* (Linne, 1758) on the East coast of South Africa. *South African Journal of Marine Science* **17**: 69-78.

Woram, R.A., McGowan, C., Stout, J.A., Gharbi, K., Ferguson, M.M., Hoyheim, B., Davidson, E.A., Davidson, W.S., Rexroad, C., Danzmann, R.G. 2004. A genetic linkage map

for Artic char (*Salvelinus alpinus*): evidence for higher recombination rates and segregation distortion in hybrid versus pure strain mapping parents. *Genome* **47**: 304-315.

Wright, J.M. 1993. DNA fingerprinting in fishes. In Hochachka, P.W., Mommsen, T. (eds). *Biochemistry and Molecular Biology of Fishes.* Elsevier, Amsterdam, pp 58-91.

Wu, S., Collins, G., Sedgley, M. 2004. A molecular linkage map of olive (*Olea europaea* L.) based on RAPD, microsatellite, and SCAR markers. *Genome* **47**: 26-35.

Young, N.D. 2000. Chapter 3: Constructing a plant genetic linkage map with DNA markers. In Phillips, R.L., Vasil, J.K (eds). *DNA-based markers in plants.* Netherlands, Kluwer Academic Publishers, pp 31-42.

Young, W.P., Wheeler, P.A., Coryell, V.H., Keim, P., Thorgaard, G.H. 1998. A detailed linkage map of rainbow trout produced using double haploids. *Genetics* **48**: 839-850.

Yu, Z., Guo, X. 2003. Genetic linkage map of the eastern oyster *Crassostrea virginica* Gmelin. *Biological Bulletin* **204**: 327-338.

Yu, Z., Guo, X. 2005. Identification and mapping of disease-resistance QTL in the eastern oyster, *Crassostrea virginica* Gmelin. *Aquaculture* **254**: 160-170.

Zeng, Z.B. 1993. Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. *Proceedings of the National Academy of Sciences* (*USA*) **90**: 10972-10976.

Zhu, C., Zhang, Y.-M. 2007. An EM algorithm for mapping segregation distortion loci. *BioMed Central Genetics* **8**: 82, doi: 10.1186/1471-2156-8-82.

Zhu, C., Wang, C., Zhang, Y.-M. 2007. Modeling segregation distortion for viability selection I. Reconstruction of linkage maps with distorted markers. *Theoretical and Applied Genetics* **114**: 295-305.

APPENDIX A

Badenhorst and Roodt-Wilding, 2007

APPENDIX A - Badenhorst and Roodt-Wilding, 2007

Aquaculture Research, 2007, 38, 546 549 doi: 10.1111/j.1365-2109.2007.01698.x

SHORT COMMUNICATION

Application of various DNA extraction methodologies on abalone, *Haliotis midae*, larvae and juveniles for fluorescent AFLP analysis

Daleen Badenhorst & Rouvay Roodt-Wilding
Department of Genetics, University of Stellenbosch, Matieland, South Africa

Correspondence: D Badenhorst, Department of Genetics, University of Stellenbosch, Private Bag X1, Matieland, 7602, South Africa. E-mail: dals@sun.ac.za

In South Africa, molecular genetic research on *Haliotis midae*, one of only six abalone species in Southern Africa, and the only species with commercial value (Cook 1998), is a fairly new field of research. To support the international market demand and as wild stocks are no longer sustainable, *H. midae* is commercially produced in hatcheries (Sales & Britz 2001). It is hoped that elucidation of the abalone genome will aid in the commercial production of abalone; by generating various markers, including for example AFLPs, genetic improvement can be mediated by marker-assisted selection (MAS) whereby superior (faster growth, better conversion efficiencies, increased disease resistance etc.) individuals can be selected at a very early age based on the marker characteristics associated with these traits (Roodt-Wilding & Slabbert 2006).

DNA suitable for molecular studies concerning abalone is commonly obtained from the epipodia (sensory tentacles) using non-destructive sampling techniques (Slabbert & Roodt-Wilding 2007, in press). These can be sampled from animals older than approximately 1 year of age (± 3–4 cm). In studies that have time constraints, alternative sources of DNA must be considered. The viability of abalone larvae (1–5 days) and juveniles (± 5 days to 1 year) are here considered as possible DNA sources for fluorescent AFLP analysis. As small amounts of tissue are involved and high-quality DNA is required for AFLPs, variations of the CTAB extraction protocol (usually used for the adult epipodia) and lysis (extraction) buffer protocols are considered.

The samples tested consisted of randomly sampled 5-day-old larvae (veliger larvae) and 2 (± 2 mm)- and 3.5 (± 5 mm)-month-old juveniles.

As only microscopic quantities of tissue are available from larvae, the CTAB extraction protocol used for adult abalone (epipodial extraction) was not appropriate. Individual larvae were sampled under a stereomicroscope, using a micropipette.

Alternative protocols were applied for the larvae: (A) the Chelex protocol as described by Walsh, Metzger and Higuchi (1991). (B) A variation of the protocol Ki, Jang and Han (2005), using 30 µL extraction buffer (autoclaved distilled water, 10 mg mL ^{1} proteinase K) and a single larva in a 0.5 mL micro-centrifuge tube. A one-step incubation/denaturation protocol was used, incubating the sample for 50 min at 55 °C followed by denaturation at 95 °C for 10 min. (C) The protocol as described by Vadopalas, Bouma, Jackels and Friedman (2006); single whole larva were incubated in a 200 µL lysis buffer [10 mM Tris/HCl (pH 8.3), 50 mM KCl, 0.5% Tween 20] and 10 mg mL ^{1} proteinase K, incubated for 2 h at 55 °C followed by 30 min at 95 °C. In addition to this, protocols as described by Simpson, Wilding and Grahame (1999), (D), and by Gruenthal and Burton (2005), (E), using an identical lysis buffer as above, were tested. Only the extraction volume and the one-step incubation/denaturation time differed: 15 µL extraction buffer, 60 min at 65 °C followed by 15 min at 94 °C (D) and 10 µL lysis buffer, 1 h at 60 °C followed by 20 min at 80 °C (E). (F) These three protocols (methods C–E) were later modified to include 50 mM EDTA as EDTA

APPENDIX A

inhibits the action of DNases. (G) The products of the extractions using a lysis buffer (methods C–E) were, in addition, precipitated with 100% ethanol after the incubation/denaturation step. The pellets were washed twice in 70% ethanol, dried and suspended in 50 µL sterile ddH_2O to determine if more concentrated genomic DNA is obtained compared with using only the lysis buffer. (H) A variation of the protocol described by Mo and Rinkevich (2001) was used, after isolation of larval DNA using the lysis buffer (from C to E). After completion of the incubation/denaturation step, DNA was extracted by equilibrated phenol (pH 8) and once with phenol–chloroform–isoamyl alcohol, followed by one chloroform–isoamyl alcohol step. DNA was precipitated with 100% ethanol and the pellet was washed twice with 70% ethanol and extracted DNA was dissolved with 40 µL sterile distilled water.

The juvenile (2- and 3.5-month old) abalone were both extracted using a variation of the protocol described by Li and Guo (2004) (based on Saghai-Maroof, Soliman, Jorgensen & Allard 1984). The 2-month-old individuals were extracted whole with their shell due to their diminutive size, whereas only tissue was used of the 3.5-month-old individuals where the tissue could be removed from their shells. DNA was extracted from these abalone tissues using the phenol–chloroform protocol (Li & Guo 2004); the tissue was placed in 700 µL lysis buffer [2% CTAB, 100 mM Tris/HCl (pH 6.8), 1.4 M NaCl, 20 mM EDTA (pH 8), 0.2% 2-β mercaptoethanol] and 5 µL 10 mg mL $^{-1}$ proteinase K, briefly vortexed and incubated at 60 °C overnight. DNA was extracted once with phenol–choloroform–isoamyl alcohol (25:24:1), and twice with cholorofom–isoamyl alcohol (24:1) and then precipitated with isopropanol (cold) overnight in the freezer. Pellets were washed twice in 70% ethanol, dried and suspended in 50 µL sterile ddH_2O and stored at 2–6 °C.

All DNA concentrations were measured using a Nanodrop spectrophotometer (ND-1000 Spectrophotometer; NanoDrop Technologies) and when concentrations were > 100 ng µL $^{-1}$ they were adjusted to 100 ng µL $^{-1}$ for AFLP analysis.

For the larvae, the Chelex extraction (method A), the distilled water protocol (method B), the additional precipitation steps (method G) and the added phenol–chloroform–isoamyl wash steps (method H) yielded the poorest results (< 8 ng µL $^{-1}$) and were subsequently not considered for AFLP analysis. The Simpson *et al.* (1999) and Gruenthal and Burton (2005) protocols with added/additional 50 mM EDTA (method F) yielded superior results (± 50–80 ng µL $^{-1}$) compared with the protocols using only lysis buffer without EDTA (± 20–40 ng µL $^{-1}$) (methods D and E) and the Vadopalas *et al.* 2006 protocol (± 40 ng µL $^{-1}$) (method C). The superior extracts (method F) and those from methods C, D and E were subsequently used for fluorescent AFLP analysis.

With the older juveniles, 70–100 and > 300 ng µL $^{-1}$ DNA were yielded with the 2 and 3.5-month-old juveniles. The 3.5-month-old samples yielded the best results and were adjusted to 100 ng µL $^{-1}$ for AFLP analysis. These extracts could be visualized through electrophoresis on an agarose gel stained with ethidium bromide unlike the larvae extracts (Fig. 1).

DNA obtained from the extraction protocols was used for fluorescent AFLP analysis. AFLP amplifications were conducted with the AFLP Plant Mapping Kit (Applied Biosystems, Foster City, CA, USA) and by analyses conducted on the ABI3130xl Genetic Analyser (Applied Biosystems, Foster City, CA, USA). The restriction enzymes *Mse*I and *Eco*RI were used. (Results summarized in Table 1)

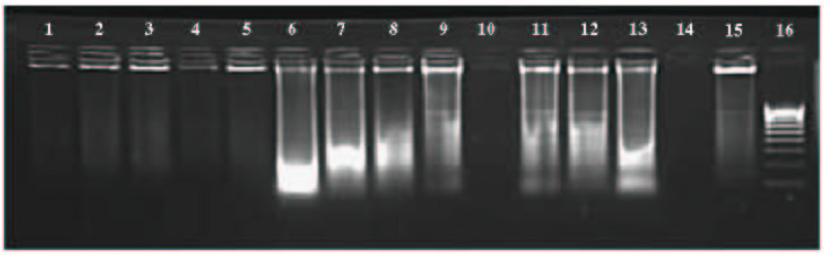


Figure 1 Agarose gel depicting the quality of the juvenile genomic DNA extractions obtained by using the protocol of Li and Guo (2004). Lanes 1–5 contain extracted DNA from 2-month-old juvenile samples, while lanes 6–15 contain the 3.5-month-old juveniles' extracted DNA samples. Lane 16 contains a 1 kb DNA Ladder (Bioline, London, UK).

APPENDIX A

Table 1 Summary of samples and DNA extraction protocols used and results of the extractions and fluorescent AFLP analysis

Samples	Protocol	Extraction type	Method	DNA extraction results	AFLP analysis	AFLP analysis results
Larvae (5 days old)	Walsh *et al.* (1991)	Chelex	A	Poor (<8 ng μL^{-1})	No	N/A
	Ki *et al.* (2005)	A Lysis buffer	B			
	Vadopalas *et al.* (2006)		C	Average (± 40 ng μL^{-1})	Yes	Poor (no visible peaks)
	Simpson *et al.* (1999)		D	Average (± 20–40 ng μL^{-1})	Yes	
	Gruenthal and Burton (2005)		E			
	Methods C–E	A Lysis buffer+added EDTA	F	Good (± 50–80 ng μL^{-1})	Yes	
	Methods C–E	A Lysis buffer+additional precipitation steps	G	Poor (<8 ng μL^{-1})	No	N/A
	Mo and Rinkevich (2001)	A Lysis buffer+wash steps (i.e. chloroform–isoamyl alcohol)	H			
2-month-old juveniles (± 2 mm)	Li and Guo (2004)	CTAB extraction		Good (± 70–100 ng μL^{-1})	Yes	Poor results
3.5-month-old juveniles (± 5 mm)	Li and Guo (2004)	CTAB extraction		Best results (> 300 ng μL^{-1})	Yes	Good

Of all the extracts tested with the AFLP Plant Mapping Kit, only the 3.5-month-old tissue samples produced results that could be reliably scored using GeneMapper 3.7, as none of the others yielded any form of peaks/alleles. According to this study, to obtain DNA of stringent and suitable quality for fluorescent AFLP analysis from the abalone tissue, the animals have to be approximately 3 months of age (± 5 mm length).

Although the extracts obtained from larvae using the mentioned lysis buffers are sufficient for polymerase chain reaction (Vadopalas *et al.* 2006), it is not suitable for fluorescent AFLP analysis. This may be due to the extracts being of poor-quality DNA resulting in the restriction–ligation reaction being impeded. The DNA obtained from larvae using these lysis buffers is of such a nature that the DNA may be partly degraded or that the DNA is not enough for AFLP analysis, as ideally 100 ng μL^{-1} is required. This quantity could probably be less for abalone tissue as a fluorescent AFLP trial run carried out using DNA obtained from adult abalone epipodia, adjusted to 20 ng μL^{-1}, gave good results.

This is the first study of its kind concerning abalone, specifically using fluorescent AFLP analysis with *H. midae*. There is still much scope for future research and improvement on the present extraction protocols for obtaining high-quality DNA from larvae and small juveniles for use in AFLP analysis and in other molecular techniques and may find application with researchers interested in techniques that enable them to conduct their research over a shorter period of time with regards to slow-growing species.

Acknowledgments

Stellenbosch University is thanked for providing research facilities, and the South African abalone industry as well as National Research Foundation are thanked for providing funding for this research. Lize Schoonbee of Irvin and Johnson abalone hatchery (Danger Point, South Africa) are thanked for supplying abalone larvae and small juveniles.

References

Cook P. (1998) The current status of abalone farming in South Africa. *Journal of Shellfish Research* **17**, 601 602.

Gruenthal K.M. & Burton R.S. (2005) Genetic diversity and species identification in the endangered white abalone (*Haliotis sorenseni*). *Conservation Genetics* **6**, 929 939.

APPENDIX A

Ki J.-S., Jang G.Y. & Han M-S. (2005) Integrated method for single-cell DNA extraction, PCR amplification, and sequencing of ribosomal DNA from harmful Dinoflagellates *Cochlodinium polykrikoides* and *Alexandrium catenella*. *Marine Biotechnology* **6**, 587 593.

Li L. & Guo X. (2004) AFLP-based genetic linkage maps of the Pacific oyster *Crassostrea gigas* Thunberg. *Marine Biotechnology* **6**, 26 36.

Mo C. & Rinkevich B. (2001) A simple, reliable, and fast protocol for Thraustochytrid DNA extraction. *Marine Biotechnology* **3**, 100 102.

Roodt-Wilding R. & Slabbert R. (2006) Molecular markers to assist the South African abalone industry. *South African Journal of Science* **102**, 99 102.

Saghai-Maroof M.A., Soliman K.M., Jorgensen R.A. & Allard R.W. (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences of the USA* **81**, 8014 1018.

Sales J. & Britz P.J. (2001) Research on abalone (*Haliotis midae* L.) cultivation in South Africa. *Aquaculture Research* **32**, 863 874.

Simpson R.J., Wilding C.S. & Grahame J. (1999) Simple protocol for extracting nuclear DNA from single embryos of a marine snail. *BioTechniques* **26**, 1050 1052.

Slabbert R. & Roodt-Wilding R. (2007) Non-destructive sampling of juvenile abalone using epipodial tentacles and mucus: method and application. *African Journal of Marine Science* **28**, 719 721.

Vadopalas B., Bouma J.V., Jackels C.R. & Friedman C.S. (2006) Application of real-time PCR for simultaneous identification and quantification of larval abalone. *Journal of Experimental Marine Biology and Ecology* **334**, 219 228.

Walsh P.S., Metzger D.A. & Higuchi R. (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-Based typing from forensic material. *BioTechniques* **10**, 506 513.

Keywords: Abalone, AFLP, DNA extraction, *Haliotis midae*, larvae, juvenile

APPENDIX B

Parent and 3.5-month-old offspring genomic DNA extractions

APPENDIX B – Parent and 3.5-month-old offspring genomic DNA extractions

Sample No.	[ng/ul]	260/280 ratio	260/230 ratio	Diluted
Female parent (F23)	273.8	1.82	1.87	Yes
Male parent (M22)	157.9	1.86	2.20	No
001	367.0	2.04	1.64	Yes
002	334.6	2.04	1.64	Yes
003	511.1	1.97	1.79	Yes
004	347.5	2.04	2.76	Yes
005	824.7	2.15	2.25	Yes
006	303.4	2.01	2.29	Yes
007	61.0	2.41	2.53	
008	174.4	2.22	1.92	No
009	357.2	1.98	2.16	Yes
010	232.6	2.03	2.20	Yes
011	539.9	1.90	1.83	Yes
012	277.6	1.99	2.20	Yes
013	248.3	1.95	2.30	Yes
014	354.8	1.98	2.00	Yes
015	185.9	1.93	2.19	No
016	188.6	1.98	2.02	No
017	364.6	1.89	1.81	Yes
018	343.1	1.91	2.10	Yes
019	726.8	2.05	2.23	Yes
020	388.7	1.96	1.91	Yes
021	269.8	1.96	2.07	Yes
022	175.2	1.93	2.20	No
023	173.3	1.94	2.08	No
024	301.6	1.95	2.19	Yes
025	380.2	1.92	2.03	Yes
026	263.8	1.92	2.21	Yes
027	87.5	1.89	2.11	
028	319.2	2.02	2.12	Yes
029	417.2	1.95	1.87	Yes
030	240.2	1.98	1.92	Yes
031	174.8	1.98	2.14	No
032	45.8	1.95	1.87	
033				
034	226.7	1.97	2.23	Yes
035	431.8	1.95	1.96	Yes
036	418.8	1.96	2.06	Yes
037	225.2	1.96	2.14	Yes
038				
039	15.4	1.86	1.84	

040	229.1	1.97	2.24	Yes
041	389.0	1.92	2.06	Yes
042	297.0	1.92	2.47	Yes
043	463.0	1.98	2.31	Yes
044	463.3	1.92	2.23	Yes
045	618.8	2.01	1.94	Yes
046	591.9	2.08	2.12	Yes
047	516.0	2.00	2.18	Yes
048	468.8	1.99	2.04	Yes
049	359.3	1.93	2.10	Yes
050	558.8	1.97	2.14	Yes
051	312.7	2.02	2.23	Yes
052	458.5	1.95	2.12	Yes
053	581.9	1.93	1.90	Yes
054	375.8	1.99	2.08	Yes
055	547.2	1.96	2.10	Yes
056	562.6	2.01	2.08	Yes
057	757.3	1.99	2.12	Yes
058	501.9	2.00	2.13	Yes
059	57.6	2.12	2.40	
060	344.2	1.91	2.20	Yes
061	238.5	1.94	2.17	Yes
062	335.0	1.96	2.17	Yes
063	371.7	1.97	2.00	Yes
064	313.6	1.94	2.28	Yes
065	307.6	2.04	2.23	Yes
066	265.3	1.99	2.23	Yes
067	208.8	2.01	2.34	Yes
068	232.9	1.96	2.29	Yes
069	339.1	2.07	2.28	Yes
070				
071	429.3	2.03	2.03	Yes
072	253.2	1.94	2.10	Yes
073	289.8	2.02	2.11	Yes
074	437.3	2.01	1.99	Yes
075	300.9	2.00	2.37	Yes
076	178.8	1.96	2.10	No
077	289.3	2.03	2.14	Yes
078	334.2	2.04	2.00	Yes
079	307.6	2.03	2.16	Yes
080				
081	370.3	1.99	2.39	Yes
082	28.9	2.08	2.16	
083	403.5	2.04	2.21	Yes
084				
085	52.3	1.85	1.44	
086	85.8	1.93	2.90	No
087	95.1	1.99	3.05	No
088	116.3	2.05	2.46	No

089				
090	163.9	1.97	2.07	No
091	79.3	2.01	2.57	No
092	181.0	1.95	2.24	No
093	88.3	2.01	2.66	No
094	196.5	1.99	2.36	Yes
095	152.7	1.94	2.04	Yes
096	309.2	1.93	2.27	Yes
097	242.0	1.99	2.26	Yes
098				
099	313.9	1.99	2.19	Yes
100	309.6	1.96	2.26	Yes
101	413.8	1.99	2.32	Yes
102	159.9	1.92	2.33	No
103	111.7	1.98	2.36	No
104	278.5	1.96	2.40	Yes
105	283.6	1.99	2.17	Yes
106	261.6	1.95	2.28	Yes
107	194.8	1.95	2.38	Yes
108	140.8	2.05	2.40	No
109	81.1	1.89	1.95	No
110	127.2	2.01	1.84	No
111	28.8	1.82	1.23	
112	332.2	1.93	1.96	Yes
113	732.2	2.08	2.27	Yes
114	88.2	1.94	1.85	No
115	97.0	1.88	1.87	No
116	431.6	1.96	2.36	Yes
117	537.7	1.94	1.99	Yes
118	471.5	1.89	1.80	Yes
119	390.2	1.94	2.29	Yes
120	542.2	1.91	2.23	Yes
121	310.5	1.85	1.77	Yes
122	408.9	1.89	1.78	Yes
123	103.2	1.9	2.05	No
124	161.1	1.96	2.03	No
125	829.9	2.00	1.96	Yes

Samples were diluted to 150ng/µl; if concentration was close to 150ng/µl the stock was used (sample was not diluted).

Colour key:

	Low concentration DNA
	DNA extraction unsuccessful
	AFLP profile different from parents

APPENDIX C

Chi-square Table

APPENDIX C – Chi-square (χ^2) critical values

	Tail probability p										
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4

Source: http://rvgs.k12.va.us/statman/.

APPENDIX D

Segregation analysis of 1:1 segregating AFLP markers in the progeny

APPENDIX D – Segregation analysis of 1:1 segregating AFLP markers in the progeny

Table D1. Polymorphic markers of the female parent segregating through the progeny.

Marker	No. of progeny	χ^2 - value (df = 1)	Probability value
A1f61	108	1.82	$P > 0.05$
A1f77	108	59.11	$P < 0.05$
A1f79	108	30.50	$P < 0.05$
A1f90	108	13.66	$P < 0.05$
A1f263	108	35.30	$P < 0.05$
A1268	108	4.51	$P > 0.05$*
A1f345	108	4.51	$P > 0.05$*
A1f422	108	24.06	$P < 0.05$
A1f480	108	4.51	$P > 0.05$*
A2f50	108	3.73	$P > 0.05$
A2f78	108	3.01	$P > 0.05$
A2f155	108	0.04	$P > 0.05$
A2f180	108	1.34	$P > 0.05$
A2f198	108	43.32	$P < 0.05$
A2f216	108	0.04	$P > 0.05$
A2f261	108	1.34	$P > 0.05$
A2f267	108	24.06	$P < 0.05$
A2f344	108	10.89	$P < 0.05$
A2f377	108	0.59	$P > 0.05$
A2f395	108	4.51	$P > 0.05$*
A2f430	108	10.89	$P < 0.05$
A2f458	108	35.30	$P < 0.05$
A2f470	108	0.15	$P > 0.05$
A2f472	108	6.32	$P > 0.05$*
E2f75	108	4.51	$P > 0.05$*
E2f95	108	15.17	$P < 0.05$
E2f113	108	0.93	$P > 0.05$
E2f125	108	1.82	$P > 0.05$
E2f158	108	0.15	$P > 0.05$
E2f165	108	28.26	$P < 0.05$
E2f270	108	0.93	$P > 0.05$
E2f407	108	13.66	$P < 0.05$
D2f60	108	35.30	$P < 0.05$
D2f85	108	28.26	$P < 0.05$
D2f103	108	0.93	$P > 0.05$
D2f107	108	6.32	$P > 0.05$*
D2f110	108	43.32	$P < 0.05$
D2f138	108	6.32	$P > 0.05$*
D2f161	108	3.73	$P > 0.05$
D2f180	108	0.93	$P > 0.05$
D2f202	108	8.44	$P > 0.05$*
D2f206	108	4.51	$P > 0.05$*

D2f238	108	5.38	$P > 0.05$*
D2f263	108	43.32	$P < 0.05$
D2f272	108	5.38	$P > 0.05$*
D2f376	108	28.26	$P < 0.05$
D6f162	107	0.76	$P > 0.05$
D6f296	107	4.98	$P > 0.05$*
C1f51	107	1.58	$P > 0.05$
C1f93	107	1.58	$P > 0.05$
C1f107	107	29.67	$P < 0.05$
C1f126	107	0.23	$P > 0.05$
C1f158	107	10.35	$P < 0.05$
C1f200	107	21.37	$P < 0.05$
B5f59	108	59.11	$P < 0.05$
B5f79	108	3.73	$P > 0.05$
B5f116	108	10.89	$P < 0.05$
B5f226	108	52.40	$P < 0.05$
B5f251	108	20.23	$P < 0.05$
B5f256	108	6.32	$P > 0.05$*
B5f263	108	0.00	$P > 0.05$
B5f265	108	15.17	$P < 0.05$
B5f358	108	10.89	$P < 0.05$
B1f83	108	30.50	$P < 0.05$
B1f115	108	7.34	$P > 0.05$*
F6f109	108	5.38	$P > 0.05$*
F6f113	108	0.00	$P > 0.05$
F6f116	108	10.89	$P < 0.05$
F6f174	108	20.23	$P < 0.05$
F6f177	108	16.77	$P < 0.05$
F6f265	108	12.23	$P < 0.05$
F6f284	108	5.38	$P > 0.05$*
B3f58	108	3.01	$P > 0.05$
B3f75	108	24.06	$P < 0.05$
B3f84	108	8.44	$P > 0.05$*
B3f100	108	28.26	$P < 0.05$
B3f105	108	20.32	$P < 0.05$
B3f115	108	40.53	$P < 0.05$
B3f118	108	46.22	$P < 0.05$
B3f138	108	16.77	$P < 0.05$
B3f143	108	3.73	$P > 0.05$
B3f151	108	1.82	$P > 0.05$
B3f162	108	5.38	$P > 0.05$*
B3f176	108	12.23	$P < 0.05$
B3f179	108	0.15	$P > 0.05$
B3f203	108	0.00	$P > 0.05$
B3f215	108	4.51	$P > 0.05$*
B3f227	108	0.59	$P > 0.05$
B3f264	108	16.77	$P < 0.05$
B3f273	108	0.04	$P > 0.05$
B3f295	108	4.51	$P > 0.05$*
B3f305	108	9.63	$P < 0.05$**
B3f351	108	10.89	$P < 0.05$

B4f119	108	15.17	$P < 0.05$
B4f123	108	7.34	$P > 0.05$*
B4f160	108	13.66	$P < 0.05$
B4f161	108	0.15	$P > 0.05$
B4f235	108	13.66	$P < 0.05$
B4f239	108	9.63	$P < 0.05$**
B4f243	108	9.63	$P < 0.05$**
B4f398	108	7.34	$P > 0.05$*
D3f74	108	49.25	$P < 0.05$
D3f140	108	18.46	$P < 0.05$
D3f256	108	7.34	$P > 0.05$*
D3f296	108	32.85	$P < 0.05$
D3f298	108	5.38	$P > 0.05$*
D3f328	108	13.66	$P < 0.05$

* Non-significant following Bonferroni correction ($P > 0.05$).
** Distorted markers ($P < 0.05$ and $P > 0.001$).

Table D2. Polymorphic markers of the male parent segregating through the progeny.

Marker	No. of progeny	χ^2 - value (df = 1)	Probability value
A1f63	108	0.148	$P > 0.05$
A1f71	108	32.85	$P < 0.05$
A1f81	108	16.772	$P < 0.05$
A1f102	108	22.097	$P < 0.05$
A1f227	108	26.107	$P < 0.05$
A1f238	108	16.772	$P < 0.05$
A1f267	108	3.725	$P > 0.05$
A2f129	108	0	$P > 0.05$
A2f311	108	4.513	$P > 0.05$*
E2f52	108	13.66	$P < 0.05$
E2f157	108	18.458	$P < 0.05$
E2f167	108	0.148	$P > 0.05$
E2f186	108	7.342	$P > 0.05$*
E2f261	108	24.055	$P < 0.05$
E2f293	108	26.107	$P < 0.05$
D2f105	108	3.725	$P > 0.05$
D2f131	108	1.336	$P > 0.05$
D2f379	108	1.336	$P > 0.05$
D6f66	107	0.757	$P > 0.05$
D6f222	107	0.757	$P > 0.05$
D6f244	107	17.778	$P < 0.05$
D6f291	107	1.583	$P > 0.05$
C1f66	107	1.583	$P > 0.05$
C1f248	107	27.447	$P < 0.05$
C1f315	107	36.958	$P < 0.05$
C1f327	107	1.132	$P > 0.05$
C1f358	107	1.583	$P > 0.05$
C1f442	107	29.669	$P < 0.05$
B5f62	108	49.245	$P < 0.05$

B5f106	108	0.593	$P > 0.05$
B5f229	108	35.505	$P < 0.05$
B5f245	108	9.625	$P < 0.05$**
B5f260	108	3.014	$P > 0.05$
B5f339	108	0.037	$P > 0.05$
B1f59	108	2.379	$P > 0.05$
B1f63	108	4.513	$P > 0.05$*
B1f75	108	13.66	$P < 0.05$
B1f107	108	0.593	$P > 0.05$
B1f205	108	37.862	$P < 0.05$
B1f266	108	35.505	$P < 0.05$
B1f373	108	32.85	$P < 0.05$
F6f81	108	4.513	$P > 0.05$*
F6f274	108	0.037	$P > 0.05$
F6f285	108	26.107	$P < 0.05$
B3f113	108	6.321	$P > 0.05$*
B3f121	108	22.097	$P < 0.05$
B3f165	108	37.862	$P < 0.05$
B3f169	108	0.593	$P > 0.05$
B3f197	108	20.232	$P < 0.05$
B3f247	108	3.014	$P > 0.05$
B3f309	108	9.625	$P < 0.05$**
B4f70	108	0.927	$P > 0.05$
B4f147	108	3.014	$P > 0.05$
B4f149	108	1.819	$P > 0.05$
B4f264	108	4.513	$P > 0.05$*
B4f311	108	26.107	$P < 0.05$
D3f57	108	15.173	$P < 0.05$
D3f62	108	3.725	$P > 0.05$
D3f125	108	35.303	$P < 0.05$
D3f176	108	28.255	$P < 0.05$
D3f186	108	13.66	$P < 0.05$
D3f204	108	4.513	$P > 0.05$*
D3f213	108	15.173	$P < 0.05$
D3f267	108	8.443	$P > 0.05$*
D3f318	108	35.303	$P < 0.05$
D3f353	108	6.321	$P > 0.05$*

* Non-significant following Bonferroni correction ($P > 0.05$).
** Distorted markers ($P < 0.05$ and $P > 0.001$).

APPENDIX E

Segregation analysis of 3:1 segregating AFLP markers

APPENDIX E – Segregation analysis of 3:1 segregating AFLP markers (present in both parents) in the progeny, used for homology detection between the two parental maps

Marker	No. of progeny	χ^2 - value (df = 1)	Probability value
A1f51	180	1.23	$P > 0.05$
A1f58	180	1.23	$P > 0.05$
A1f66	180	0.79	$P > 0.05$
A1f67	180	2.42	$P > 0.05$
A1f100	180	5.98	$P > 0.05$*
A1f165	180	1.23	$P > 0.05$
A1f177	180	0.20	$P > 0.05$
A1f203	180	5.98	$P > 0.05$*
A1f208	180	0.79	$P > 0.05$
A2f55	180	3.16	$P > 0.05$
A2f158	180	5.98	$P > 0.05$*
A2f272	180	4.68	$P > 0.05$*
A2f441	180	0.20	$P > 0.05$
B3f54	180	4.94	$P > 0.05$*
B3f57	180	4.00	$P > 0.05$*
B3f62	180	5.98	$P > 0.05$*
B3f64	180	5.98	$P > 0.05$*
B3f72	180	4.00	$P > 0.05$*
B3f78	180	4.94	$P > 0.05$*
B3f86	180	1.78	$P > 0.05$
B3f88	180	0.20	$P > 0.05$
B3f91	180	0.44	$P > 0.05$
B3f92	180	4.94	$P > 0.05$*
B3f96	180	1.78	$P > 0.05$
B3f110	180	0.79	$P > 0.05$
B3f120	180	4.94	$P > 0.05$*
B3f135	180	4.94	$P > 0.05$*
B3f156	180	1.23	$P > 0.05$
B3f177	180	4.00	$P > 0.05$*
B3f185	180	0.44	$P > 0.05$
B3f192	180	3.16	$P > 0.05$
B3f199	180	4.00	$P > 0.05$*
B3f202	180	4.94	$P > 0.05$*
B3f211	180	1.23	$P > 0.05$
B3f238	180	0.44	$P > 0.05$
B3f254	180	5.98	$P > 0.05$*
B3f286	180	3.16	$P > 0.05$
B3f291	180	0.00	$P > 0.05$
B3f301	180	1.78	$P > 0.05$
B3f306	180	4.00	$P > 0.05$*

B3f313	180	5.98	$P > 0.05$*
B4f63	180	3.16	$P > 0.05$
B4f77	180	5.98	$P > 0.05$*
B4f90	180	2.42	$P > 0.05$
B4f91	180	2.42	$P > 0.05$
B4f112	180	5.98	$P > 0.05$*
B4f125	180	1.23	$P > 0.05$
B4f174	180	3.16	$P > 0.05$
B4f183	180	1.23	$P > 0.05$
B4f203	180	2.42	$P > 0.05$
B4f216	180	3.16	$P > 0.05$
B4f226	180	0.44	$P > 0.05$
B4f232	180	4.94	$P > 0.05$*
B4f272	180	1.78	$P > 0.05$
B4f278	180	1.78	$P > 0.05$
B4f288	180	1.78	$P > 0.05$
B4f302	180	4.00	$P > 0.05$*
B4f326	180	4.00	$P > 0.05$*
B4f437	180	1.23	$P > 0.05$
B1f72	180	1.23	$P > 0.05$
B1f90	180	1.23	$P > 0.05$
B1f185	180	2.42	$P > 0.05$
B5f53	180	0.05	$P > 0.05$
B5f54	180	0.44	$P > 0.05$
B5f61	180	5.98	$P > 0.05$*
B5f67	180	3.16	$P > 0.05$
B5f69	180	1.78	$P > 0.05$
B5f152	180	1.78	$P > 0.05$
B5f167	180	4.94	$P > 0.05$*
B5f182	180	1.78	$P > 0.05$
B5f451	180	1.23	$P > 0.05$
C1f54	180	0.05	$P > 0.05$
C1f139	180	1.78	$P > 0.05$
C1f344	180	2.42	$P > 0.05$
D6f61	180	1.23	$P > 0.05$
D6f152	180	1.23	$P > 0.05$
D6f165	180	2.42	$P > 0.05$
D6f168	180	0.00	$P > 0.05$
D6f212	180	0.79	$P > 0.05$
D6f248	180	4.94	$P > 0.05$*
D2f221	180	5.98	$P > 0.05$*
D2f310	180	2.42	$P > 0.05$
D3f63	180	3.16	$P > 0.05$
D3f71	180	4.00	$P > 0.05$*
D3f75	180	4.94	$P > 0.05$*
D3f93	180	0.20	$P > 0.05$
D3f97	180	2.42	$P > 0.05$
D3f103	180	0.79	$P > 0.05$
D3f122	180	2.42	$P > 0.05$

D3f135	180	1.78	$P > 0.05$
D3f136	180	4.00	$P > 0.05$*
D3f164	180	1.78	$P > 0.05$
D3f211	180	4.94	$P > 0.05$*
D3f240	180	0.79	$P > 0.05$
D3f250	180	1.78	$P > 0.05$
D3f275	180	0.44	$P > 0.05$
D3f441	180	5.98	$P > 0.05$*
E2f59	180	3.16	$P > 0.05$
E2f63	180	4.00	$P > 0.05$*
E2f69	180	5.98	$P > 0.05$*
E2f85	180	1.78	$P > 0.05$
E2f86	180	4.00	$P > 0.05$*
E2f92	180	3.15	$P > 0.05$
E2f104	180	5.98	$P > 0.05$*
E2f106	180	0.44	$P > 0.05$
E2f114	180	3.16	$P > 0.05$
E2f122	180	0.44	$P > 0.05$
E2f147	180	4.00	$P > 0.05$*
E2f173	180	4.00	$P > 0.05$*
E2f180	180	0.05	$P > 0.05$
E2f219	180	3.16	$P > 0.05$
E2f250	180	0.44	$P > 0.05$
E2f272	180	1.23	$P > 0.05$
F6f90	180	5.98	$P > 0.05$*
F6f95	180	2.42	$P > 0.05$
F6f96	180	4.94	$P > 0.05$*
F6f104	180	0.05	$P > 0.05$
F6f120	180	1.78	$P > 0.05$
F6f129	180	5.98	$P > 0.05$*
F6f146	180	5.98	$P > 0.05$*
F6f149	180	1.78	$P > 0.05$
F6f159	180	4.00	$P > 0.05$*
F6f163	180	3.16	$P > 0.05$
F6f171	180	3.16	$P > 0.05$
F6f190	180	3.16	$P > 0.05$
F6f254	180	2.42	$P > 0.05$
F6f293	180	2.42	$P > 0.05$
F6f356	180	0.79	$P > 0.05$
A1f86	180	8.35	$P > 0.05$*
A1f137	180	8.35	$P > 0.05$*
A1f208	180	0.79	$P > 0.05$
D6f131	180	7.11	$P > 0.05$*
D6f137	180	8.35	$P > 0.05$*
D6f170	180	8.35	$P > 0.05$*
D6f191	180	7.11	$P > 0.05$*
D6f272	180	8.35	$P > 0.05$*
B3f52	180	8.35	$P > 0.05$*
B3f66	180	8.35	$P > 0.05$*

B3f116	180	7.11	$P > 0.05$*
B3f145	180	7.11	$P > 0.05$*
B3f196	180	8.35	$P > 0.05$*
B3f251	180	8.35	$P > 0.05$*
B4f133	180	7.11	$P > 0.05$*
B4f143	180	8.35	$P > 0.05$*
B4f168	180	7.11	$P > 0.05$*
B4f241	180	8.35	$P > 0.05$*
B4f256	180	8.35	$P > 0.05$*
B5f70	180	7.11	$P > 0.05$*
B5f120	180	8.35	$P > 0.05$*
B5f127	180	8.35	$P > 0.05$*
B5f130	180	7.11	$P > 0.05$*
B5f212	180	8.35	$P > 0.05$*
B5f242	180	7.11	$P > 0.05$*
C1f58	180	7.11	$P > 0.05$*
D3f54	180	8.35	$P > 0.05$*
D3f151	180	7.11	$P > 0.05$*
E2f67	180	8.35	$P > 0.05$*
E2f132	180	7.11	$P > 0.05$*
E2f144	180	7.11	$P > 0.05$*
F6f98	180	7.11	$P > 0.05$*
F6f141	180	7.11	$P > 0.05$*
F6f155	180	7.11	$P > 0.05$*

* Non-significant following Bonferroni correction ($P > 0.05$).

APPENDIX F

Microsatellite marker information (Slabbert *et al*., [in press])

APPENDIX F – Microsatellite marker information (Slabbert *et al.*, [in press])

Name	Repeat	Size (bp)	Forward (F) and Reverse (R) Primer	Label	Allele nr	Size Range (bp)	Accession No.
HmNR106D	TG_{15}	345	**F:** 5' TCC TTG GCC AGA ATA ACC 3'	**Fam**	16	329-389	DQ825709
			R: 5' TAT ATG GTC TGC ATC GCT G 3'				
HmNR136D	CA_{11}	254	**F:** 5' GAG TAA TAT GGG CAC CTC G 3'	**Vic**	20	211-309	DQ825710
			R: 5' GTT TGG AAT GTC TGA TTG GA 3'				
HmNR20M	$(TCC)_5(TAC)_7$	229	**F:** 5' CTA CAA CAA ACG CCG ATG 3'	**Fam**	11	187-289	EF063097
			R: 5' TGC AGT AAT AGG GGT ACC AG 3'				
HmNR54H	$(TTAGGG)_4$	359	**F:** 5' TAA CAC TAA GTC CCT CAC CC 3'	**Vic**	10	329-407	EF063103
			R: 5' CAT TCT ACA TTC GAC ATT CG 3'				
HmNR120T	$(TGAG)_{23}$	304	**F:** 5' TTG AGC ATG AGT CGT TGA GC 3'	**Pet**	24	235-347	EF121745
			R: 5' ACC TGC TCT TTA GCT CAG ATG G 3'				
HmNR185D	$(GT)_{13}$	137	**F:** 5' TAG AGT TCA TGT GTG TAC GTG TGC 3'	**Fam**	11	132-160	EF121750
			R: 5' TAC CTG TAA CGC GCT TGC T 3'				
HmNR180D	$(GT)_{24}$	287	**F:** 5' ACA AGG AGG CGT GAA ATC TGC 3'	**Vic**	12	269-297	EF121748
			R: 5' GCA TTG TTA CCC CCT ACA AAG ACC 3'				
HmNR258R	$(CAA)_{11}$	250	**F:** 5' GCA TCG CCT GAT TTG ATT C 3'		6	239-257	EF512272
			R: 5' CAG AAG GGT GGG TTG TAG TAT G 3'	**Pet**			
HmNR281P	$(CTCAA)_{24}$	367	**F:** 5' AAC CTT CAG TAA CCC ATG C 3'		21	225-375	EF512274
			R: 5' TGA ATA GGC ACC ATA AAG GG 3'	**Fam**			
HmNR289P	$(GTTGT)_5$	305	**F:** 5' GCA AGA CAG ACA TCC AAG AC 3'	**Pet**	4	301-316	EF512275
			R: 5' TAC AAA TCC CGA CAC AAG AG 3'				

APPENDIX G

Maternal LOD tables

APPENDIX G – LOD tables of maternal linkage groups generated by MAPMAKER 3.0. For each pair of markers compared, the bottom number is the LOD score and the top number represents the centiMorgan distance

Linkage group 1

	F6f284r	B5f263		B1f115		D2f272r		A1f61		
	E2f158r		C1f158r		C1f126r		A1f345			
A1f480r										
E2f158r	23.9 7.67									
B5f263	40.2 2.66	28.7 5.67								
C1f158r	–	34.5 3.86	53.0 1.08							
B1f115	63.2 0.52	52.0 1.17	–	29.1 5.50						
C1f126r	–	50.5 1.28	–	23.0 8.03	23.0 8.03					
D2f272r	–	56.9 0.81	–	53.0 1.08	52.0 1.17	35.9 3.50				
A1f345	–	–	–	39.1 2.84	45.6 1.83	31.6 4.64	49.8 1.37			
A1f61	–	–	–	48.4 1.50	60.1 0.65	55.5 0.90	–	28.7 5.67		
A1f480r	56.9 0.81	52.0 1.17	–	29.1 5.50	37.1 3.30	23.0 8.03	–	27.5 6.14	25.0 7.13	
A1268r	–	–	–	39.1 2.84	–	37.6 3.16	–	37.1 3.30	28.7 5.67	16.2 12.09
B3f203r	–	–	–	55.5 0.90	–	53.0 1.08	63.2 0.52	49.8 1.37	49.8 1.37	43.8 2.09
B3f151	–	–	–	55.5 0.90	54.4 0.98	40.7 2.53	54.4 0.98	47.5 1.59	–	–
B3f305	–	–	–	–	–	–	–	–	–	–

	A1268r	B3f151
	B3f203r	
B3f203r	38.6 2.97	
B3f151	–	–
B3f305	–	– 42.0 2.36

Linkage group 2

	HmNR54H	D3f298		
		D3f256		B4f123
D3f256	43.8 2.09			
D3f298	52.0 1.17	34.0 4.01		
B4f123	63.2 0.52	52.0 1.17	26.3 6.62	
A2f470	-	63.2 0.52	56.9 0.81	43.8 2.09

Linkage group 3

	B3f215r	HmNR281P		B4f161	
		D2f107r	B3f227		
D2f107r	40.2 2.66				
HmNR281P	-	17.5 11.16			
B3f227	-	31.4 4.80	29.1 5.50		
B4f161	-	-	-	34.0 4.01	
HmNR120r	63.2 0.52	52.0 1.17	-	56.9 0.81	13.3 14.42

Linkage group 4

	D2f103	A2f180	A2f216		D2f180	
	E2f113r	D2f161		HmNR106r		
E2f113r	-					
A2f180	42.0 2.36	40.2 2.66				
D2f161	60.1 0.65	34.0 4.01	26.3 6.62			
A2f216	63.2 0.52	38.6 2.97	38.6 2.97	30.1 5.22		
HmNR106r	54.4 0.98	34.0 4.01	40.2 2.66	34.0 4.01	2.8 26.56	
D2f180	52.0 1.17	38.6 2.97	42.0 2.36	32.7 4.39	7.5 20.13	4.6 23.72
E2f75	-	54.4 0.98	60.1 0.65	-	26.3 6.62	25.0 7.13	23.9 7.67

Linkage group 5

	B3f273	D6f296 B4f398	B4f243
B4f398	38.6 2.97		
D6f296	-	38.6 2.97	
B4f243	-	37.1 3.30	60.1 0.65
A2f155	-	-	- 27.5 6.14

Linkage group 6

	B3f295	B3f143
B3f143	35.6 3.64	
D2f206r	45.6 1.83	40.2 2.66

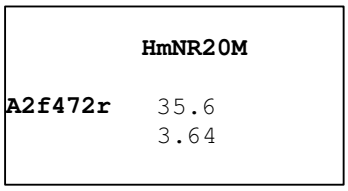
Linkage group 7

	E2f270	C1f51r
C1f51r	33.1 4.24	
C1f93r	55.5 0.90	40.7 2.53

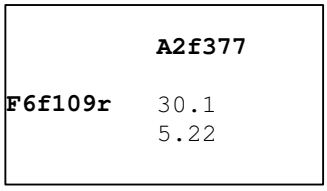
Linkage group 8

	A2f261r D2f202	B3f58	B3f179r
D2f202	63.2 0.52		
B3f58	43.8 2.09	17.4 11.38	
B3f179r	60.1 0.65	45.6 1.83	35.6 3.64
B3f273r	-	-	- 38.6 2.97

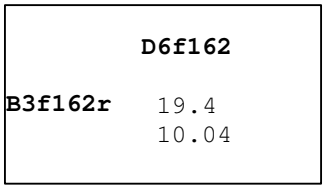
Linkage group 9



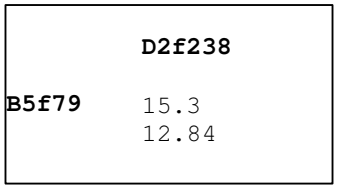
Linkage group 10



Linkage group 11



Linkage group 12



APPENDIX H

Paternal LOD tables

APPENDIX H – LOD tables of paternal linkage groups generated by MAPMAKER 3.0. For each pair of markers compared, the bottom number is the LOD score and the top number represents the centiMorgan distance

Linkage group 1

	B5f245	D3f204	B5f260	D3f353
B5f260	38.6 2.97			
D3f204	60.1 0.65	40.2 2.66		
D3f353	–	47.5 1.59	23.9 7.67	
E2f167	–	60.1 0.65	35.6 3.64	32.7 4.39

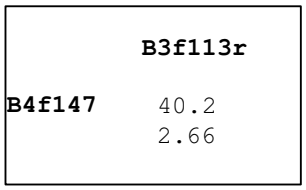
Linkage group 2

	A1f267	C1f327	C1f66	B1f59	C1f358
C1f66	27.7 5.96				
C1f327	23.0 8.03	11.4 15.90			
C1f358	26.5 6.44	14.4 13.38	10.4 16.82		
B1f59	40.2 2.66	29.1 5.50	21.9 8.60	25.4 6.95	
E2f186r	–	37.6 3.16	53.0 1.08	50.5 1.28	–

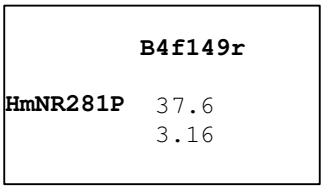
Linkage group 3

	B4f70r	D6f66r
D6f66r	43.8 2.09	
HmNR180D	63.2 0.52	17.4 11.38

Linkage group 4

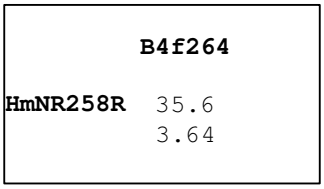


Linkage group 5

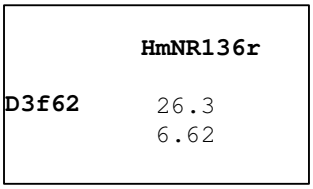


Linkage group 6

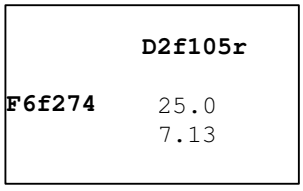
Linkage group 7



Linkage group 8



Linkage group 9



Linkage group 10